

THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

La reconnaissance vocale et l'aide aux handicapés

Thiry, Bernard

Award date:
1988

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

FACULTÉS
UNIVERSITAIRES
N. D. DE LA PAIX
NAMUR



INSTITUT D'INFORMATIQUE

Année académique
1987 - 1988

**La reconnaissance vocale
et
l'aide aux handicapés**

Bernard Thiry

Mémoire présenté en vue de
l'obtention du titre de
Licencié et Maître en Informatique

Promoteur : Jean Ramaekers

RESUME

La technologie vocale fait actuellement l'objet de développements importants dans divers domaines. Ce mémoire présente les quatres grandes orientations de cette technologie : la synthèse vocale, l'enregistrement et la digitalisation de la voix, la reconnaissance des sons et la compréhension de la parole, la reconnaissance du locuteur. Il expose plus particulièrement les caractéristiques historiques et commerciales de la reconnaissance vocale ainsi que l'avancée de la recherche dans ce domaine. Ensuite, il évalue deux produits commercialisés : le clavier de reconnaissance vocale KB 5152 V de Key Tronic et le système VoiceScribe 1000 de Dragon Systems. Enfin, il présente la reconnaissance vocale dans le cadre de son intégration à l'aide informatique aux personnes handicapées.

ABSTRACT

The Voice technology is presently experiencing a rapid growth in several applications. Within the area of man-machine communication by voice, the memoir gives information about the four main avenues of research : speech synthesis, speech store and forward, speech recognition and speaker recognition. It especially describes historical and commercial features of the speech recognition, considers major applications of voice technology (state of the art), evaluates two commercial products : the Key Tronic speech recognition keyboard model KB 5152 V and the VoiceScribe 1000 Speech Recognition System developed by Dragon Systems, and finally integrates voice into computer-based aid systems for disabled.

Je tiens à remercier :

monsieur Jean Ramaeckers, directeur de l'institut d'informatique de Namur, pour les conseils apportés au cours de la réalisation de ce travail;

le docteur R. Damper du département d'électronique de l'université de Southampton, Hampshire, Angleterre et le docteur R. M. Stephens du département de Biologie, Portsmouth Polytechnic, Hampshire, Angleterre, pour les renseignements recueillis sur l'aide informatique aux handicapés en Grande-Bretagne;

monsieur Ernest Debatty, directeur du groupe social de A., sans la générosité et la compréhension de qui, ce mémoire n'aurait pu mener à bien son étude pratique de la reconnaissance vocale;

madame Cécile Stasse pour ses remarques et son aide dans la rédaction de ce mémoire;

l'institut "De Bijtjes - Les Petites Abeilles" de Vlezenbeek, Belgique et plus particulièrement ses résidents, sa directrice et monsieur Vanliefeland Michel, logopède et responsable du pool informatique de l'institut ainsi que monsieur Christian Meunier pour leurs remarques pertinentes et leur chaleureux accueil;

monsieur Veli Pekka Rati pour son aide dans la recherche de références bibliographiques importantes;

Muriel Chandelon, Philippe Lambert, Georges Mauguit, Geneviève Warnant, pour leur lecture et leur conseils judicieux quant à la forme du mémoire;

mes parents et particulièrement Marie Eliaers, ma fiancée, pour le soutien moral qu'ils m'ont procuré et leur compréhension face au stress du mémorant;

toutes les personnes de l'Institut d'Informatique, ou extérieures à la faculté, qui ont permis de mener à bien cette étude.

- Table des matières -

Introduction	1
I. La technologie vocale	4
1.1 Introduction	5
1.2 Historique général	7
1.3 Les grandes orientations	10
1.3.1 La synthèse vocale	11
1.3.2 L'enregistrement et la digitalisation de la voix	13
1.3.3 La reconnaissance des sons et la compréhension de la parole	14
1.3.4 La reconnaissance du locuteur	17
1.4 Tableau récapitulatif	19
II. La reconnaissance vocale	20
2.1 Introduction	21
2.2 Les raisons d'utilisation	22
2.3 Les grandes étapes	24
2.4 Description de la reconnaissance vocale	30
2.4.1 Des caractéristiques liées à la parole	30
2.4.2 Des caractéristiques liées à la reconnaissance des formes	33
2.4.3 Les méthodes d'analyse et de recherche	35
2.4.4 Typologie des systèmes de reconnaissance vocale	40
2.4.5 Les qualités de l'entrée vocale	44
2.5 L'état, l'avancée de la technologie (State of the art)	50
2.5.1 Applications du secteur industriel et commercial	51
2.5.2 Applications télématiques et en télécommunications	53
2.5.3 Applications administratives et bureautiques	55
2.5.4 Applications bio-médicales et en ingénierie médicale	57
2.5.5 Applications dans l'enseignement et en informatique	60
2.5.6 Applications dans les transports	61
2.5.7 Applications "grand public"	62
2.5.8 Applications militaires	63
2.5.9 Applications diverses	64
2.6 Le marché, l'aspect commercial	65
2.7 En résumé	66

- Table des matières -

III. L'utilisation et l'expérimentation	67
3.1 Introduction	68
3.2 Key Tronic : KB 5152V	72
3.2.1 Description du matériel	72
3.2.2 Remarques générales	75
3.2.3 Comportement à l'exécution	76
3.2.4 Particularités du système	77
3.2.5 Conclusions	81
3.3 Dragon : VoiceScribe 1000	82
3.3.1 Description du matériel	82
3.3.2 Remarques générales	86
3.3.3 Comportement à l'exécution	86
3.3.4 Particularités du système	88
3.3.5 Conclusions	96
3.4 En résumé	96
IV. Le cas spécifique de l'aide aux handicapés	97
4.1 Introduction	98
4.2 Les handicapés et l'informatique	99
4.3 Les handicapés et la reconnaissance vocale	103
4.3.1 Une mise au point	103
4.3.2 Des témoignages	105
4.4 Le projet Aidami	110
4.4.1 Les fonctionnalités du système	111
4.4.2 L'architecture physique et logicielle du système	113
4.4.3 Remarques et orientations futures	116
4.5 L'intégration de la voix au projet Aidami	119
4.5.1 Première étape : destruction	121
4.5.2 Deuxième étape : inventaire	121
4.5.3 Troisième étape : construction	122
4.5.4 Quatrième étape : synthèse	123
4.6 Perspectives d'avenir	124
4.7 En résumé	127
Conclusions	128

- Table des figures et des tableaux -

Figure 1.1 : Le vocodeur de Dudley. (p. 9)

Figure 1.2 : Modèle simplifié de la reconnaissance vocale. (p. 14)

Figure 1.3 : Description d'un système de reconnaissance et de compréhension. (p. 16)

Figure 1.4 : Modèle simplifié de la vérification du locuteur. (p. 18)

Figure 1.5 : La technologie vocale, ses orientations. (p. 19)

Figure 2.1 : Représentation des sous-systèmes d'un système de reconnaissance de formes. (p. 34)

Tableau 2.2 : Récapitulatif des domaines d'applicabilité en reconnaissance vocale des techniques globale et analytique selon les paramètres d'un système de reconnaissance. (p. 38)

Tableau 2.3 : Récapitulatif des domaines d'applicabilité en reconnaissance vocale des techniques globale et analytique selon le type d'élocution des systèmes de reconnaissance. (p. 38)

Tableau 2.4 : Pourcentages moyens pour l'introduction et la correction de données via un éditeur "classique" et vocal. (p. 47)

Tableau 2.5 : Références d'applications de la reconnaissance vocale dans l'aide aux handicapés physiques. (p. 60)

Tableau 2.6 : Le marché mondial. (p. 65)

Tableau 2.7 : Le marché américain. (p. 65)

Tableau 2.8 : Le marché de la reconnaissance vocale. (p. 65)

- Table des figures et des tableaux -

Figure 3.1 : Représentation du système KB 5152 V. (p. 73)

Figure 3.2 : Représentation du système VoiceScribe
1000. (p. 91)

Figure 3.3 : Les trois types d'applications vocales avec
VS-1000. (p. 95)

Figure 4.1 : Les fonctionnalités du système Aidami. (p. 112)

Figure 4.2 : Principe du menu déroulant comme implémenté dans
la version de base de Aidami. (p. 114)

Figure 4.3 : Récapitulatif des versions du projet
Aidami. (p. 118)

Figure 4.4 : Optique de construction de Aidami et de
Aidami X. (p. 125)

Introduction

Introduction

"Le langage est une technique de communication dont l'essentiel est un système de gestes phoniques."

[POHL, 1968]

Chacun en conviendra, l'informatique est avant tout un outil, un outil performant d'aide à toutes les personnes.

Les qualités de l'informatique dispensées dans les domaines industriel, économique et scientifique ne sont plus à souligner.

Comme en témoignent les termes qui ont envahi notre vocabulaire (CFAO, CAO, EAO, bureautique, télématique, ludotique, domotique), l'informatique a envahi tous les secteurs de l'activité humaine. Nous la retrouvons partout : au travail, dans l'administration, dans les transports et jusque dans les foyers.

Mais l'informatique dans sa vocation d'aide aux personnes se découvre une autre dimension dans laquelle tous ces aspects fusionnent : une dimension sociale dans l'aide aux personnes handicapées.

Bien sûr, l'informatique ne peut guérir une infirmité mais elle peut l'alléger...

L'objectif de ce mémoire a été de nous confronter à la reconnaissance vocale de manière pratique, c'est-à-dire en tant qu'utilisateur. Dans cette perspective, la visée majeure fut de vérifier si la technologie vocale ramenée à la reconnaissance vocale pouvait représenter une aide substantielle palliant les difficultés spécifiques des handicapés souffrant d'une déficience du système sensori-moteur.

Pour ce faire, nous avons cerné au plus près la reconnaissance vocale en la décrivant, en la situant dans l'ensemble de la réalité informatique, en assurant l'apport d'un maximum d'informations notamment d'ordre bibliographique. Nous nous sommes enfin attachés à l'étude de deux systèmes de reconnaissance vocale disponibles sur le

marché; ici encore, notre souci fut d'évaluer l'adaptation de ces interfaces sur un *Personal Computer* dans l'optique d'un apport possible à l'aide aux handicapés.

Le premier chapitre situe la reconnaissance vocale au sein de la technologie vocale dont il présente l'historique, les différentes composantes et leurs articulations.

Sur cette base, le chapitre deux se consacre à la présentation de la reconnaissance vocale. Après l'avoir définie, il expose ses aspects historiques et commerciaux; l'avancée de la recherche en ce domaine est évaluée ainsi que les qualités des interfaces de ce type.

Ainsi présentée et définie la reconnaissance vocale est ensuite envisagée dans ses modalités d'application. Grâce à la bonne volonté d'un importateur et avec l'aide d'un mécène, nous avons pu disposer de deux systèmes de reconnaissance que nous avons évalués; cette phase d'évaluation fut aussi facilitée par les demandes de renseignements que nous avons précédemment adressées aux constructeurs (Chapitre III).

Ensuite, la participation à un congrès, des visites en institutions chargées de l'aide aux handicapés, les contacts répétés avec ceux-ci et leur personnel médical ont permis une juste évaluation des compétences pratiques de la reconnaissance vocale (Chapitre IV). Enfin, toujours dans ce chapitre, l'essai d'intégration d'une telle interface dans le système Aidami nous a permis de constituer d'une solution informatique visant à rencontrer les nécessités des personnes handicapées. Dans son état actuel, cette solution se présente sous une forme primitive que nous souhaitons voir améliorer par des études ultérieures auxquelles ce mémoire semble heureusement ouvrir. Nous espérons ainsi avoir posé une reconnaissance préalable à des travaux qui assureront aux personnes handicapées une aide optimale.

C H A P I T R E I .

La technologie vocale

1. Introduction
2. Historique général
3. Les grandes orientations
 1. La synthèse vocale
 2. L'enregistrement et la digitalisation de la voix
 3. La reconnaissance des sons et compréhension de la parole
 4. La reconnaissance du locuteur
4. Tableau récapitulatif

I. La technologie vocale

1.1 Introduction

"Le Robot Parlant constituait un tour de force, un dispositif dénué de toute utilité pratique et n'ayant qu'une valeur publicitaire. Une fois par heure, un groupe dirigé par un guide s'arrêtait devant lui et posait des questions.

...

Mais cela n'avait rien de bien passionnant. Il peut être utile de savoir que le carré de quatorze est cent quatre-vingt-seize, que la température ambiante est de 72 degrés Fahrenheit, que la pression de l'air est de 76 cm de mercure, ... , mais on n'a pas besoin d'un robot pour cela. Surtout il n'est pas nécessaire de posséder une masse immobile et peu maniable de fils et d'enroulements s'étendant sur vingt-cinq mètres carrés pour obtenir ce résultat.

...

-S'il vous plaît, monsieur le Robot, seriez-vous le Robot Parlant ?

On entendit un bruit d'engrenages bien huilés et une voix au timbre mécanique fit entendre des mots qui manquaient à la fois d'accent et d'intonation.

-Je... suis... le... robot... qui... parle."

[Asimov, 1950, "Robbie"]

Très souvent, le public perçoit la technologie vocale comme une invention de science-fiction. Il est vrai que de nombreux auteurs ont décrit dans les moindres détails des machines, le plus souvent des robots, douées d'un don oratoire digne des plus grands sénateurs romains (HAL 9000 de Clarke dans "2001 a space odyssey"). Aujourd'hui, cette fiction fait place à une réalité plus tangible et divergente par rapport à l'imaginaire [McCauley, 1984, pp. 131-166], [Schmandt, 1985, pp 133-159]. Mais, des confusions malheureuses sont très souvent mises en évidence dans le discours des personnes confrontées à la technologie vocale.

Les processus de reconnaissance et de synthèse vocale sont souvent pris l'un pour l'autre. Bien que les expressions "reconnaissance vocale" et "compréhension de la parole" possèdent une sémantique différente, elles sont souvent utilisées comme synonymes. Ce premier chapitre pose les bases nécessaires pour appréhender correctement les notions développées dans les chapitres suivants. Pour des raisons littéraires, nous utiliserons reconnaissance vocale et reconnaissance de la parole comme synonymes.

Les orientations principales de la technologie vocale sont la synthèse vocale, l'enregistrement et la digitalisation de la voix, la reconnaissance vocale et la reconnaissance du locuteur.

Les prochaines sections présentent un historique général de la technologie vocale et une introduction rapide à ses quatre orientations.

1.2 Historique général

Les hommes ont toujours été animés du désir de réaliser des dispositifs artificiels reproduisant mécaniquement leurs propres actions ou celles des animaux qu'ils avaient domestiqués.

Dans l'antiquité, d'habiles artisans fabriquèrent des figurines animées manuellement ou par des procédés pneumatiques et hydrauliques; certains dispositifs reproduisaient par exemple le chant des oiseaux.

A l'époque de la Renaissance, les jardins des personnes bien nommées étaient agrémentés de buissons artificiels dans lesquels sifflaient et s'ébrouaient de petits automates-oiseaux actionnés par les fontaines proches.

Durant le dix-huitième siècle, les constructeurs s'attachent à réaliser des automates androïdes dont la perfection mécanique ne sera plus égalée. Ainsi, le célèbre joueur de flûte et de galoubet du mécanicien français Jacques de Vaucanson, dont on a complètement perdu la trace aujourd'hui ! D'autres noms prestigieux sont encore à retenir comme ceux des constructeurs suisses Jaquet-Droz et Leschot. A la fin de ce siècle apparurent les premières machines mécaniques produisant des sons proches de la voix humaine.

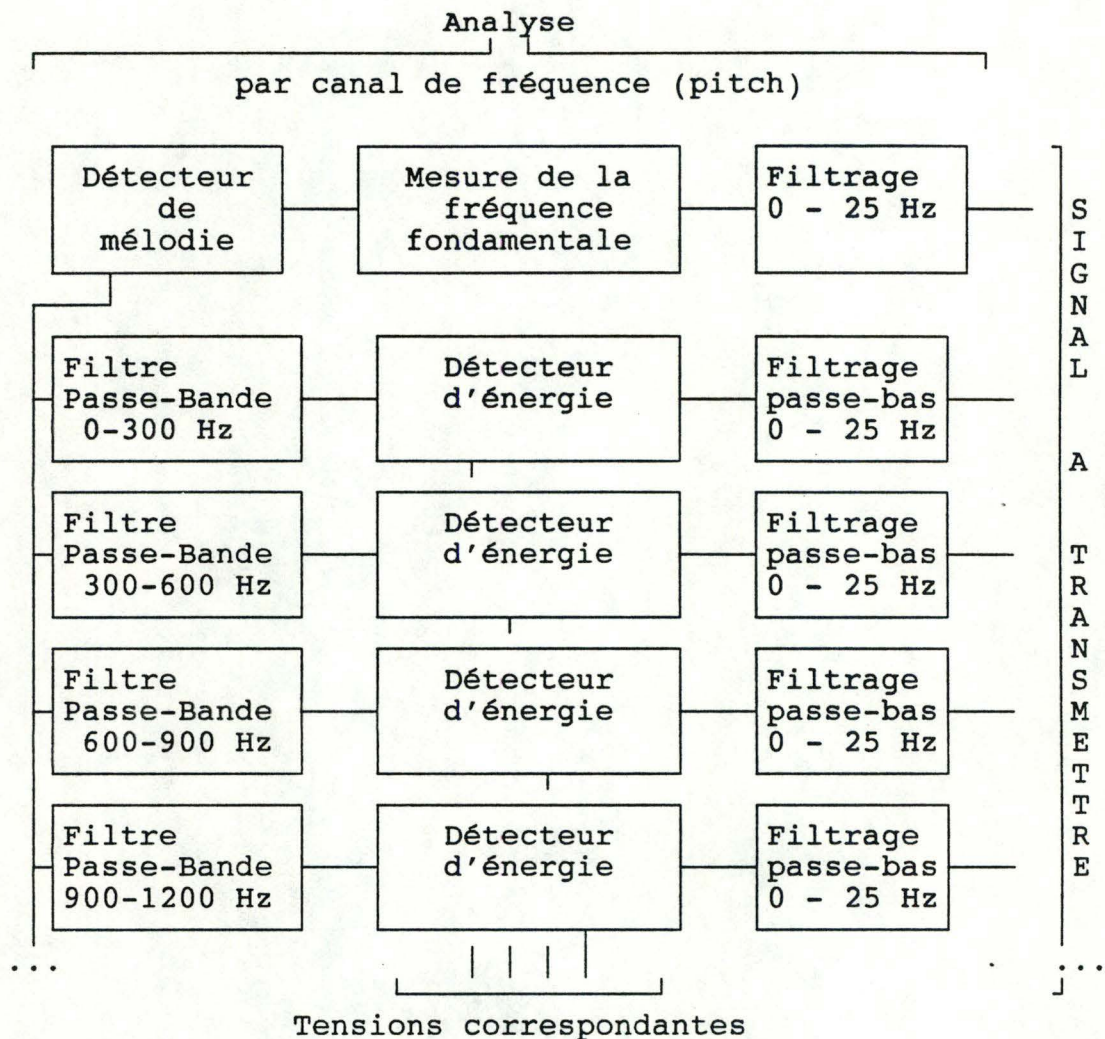
En 1779, l'Académie impériale de Saint-Petersbourg lance un défi : "construire une machine prononçant les sons des voyelles". Christian Gotlieb Kratzenstein construit une série de résonateurs acoustiques de dimensions et de formes similaires à celles de la bouche humaine.

En 1790, le baron Wolfgang von Kempflen, très connu pour son Joueur d'Echecs, qui dissimulait en fait un officier polonais, édifie une autre machine à parler reproduisant les voyelles et quelques consonnes.

Au 19ème siècle, Joseph Faber présente une machine chantant le "God Save the Queen". Même Graham Bell s'est essayé à la construction d'une machine parlante. Construite

avec la collaboration de son frère Melville, elle est basée sur le moulage d'un crâne humain et émet des voyelles, des nasales et même des mots complets.

Il faudra attendre 1939 avant de voir apparaître de réelles innovations par rapport à la machine de Faber. A l'exposition internationale de New-York, Homer Dudley présente un appareil nommé VODER - "Voice operation demonstrator". Cet appareil consiste en un synthétiseur électrique composé d'un oscillateur à relaxation pour la production de sons voisés et d'un générateur de bruits blancs pour les autres sons. Ce même Dudley présente aussi une machine ayant pour but de faciliter les recherches menées aux laboratoires de Bell Telephone au sujet de la réduction de la largeur de la bande passante nécessaire aux conversations téléphoniques, en vue de transmettre la voix avec une plus grande efficacité. Cette machine VOCODER - "Voice Coder" - décompose la voix en dix signaux que l'on peut utiliser pour commander un synthétiseur similaire à celui du Voder. Vous pouvez voir le schéma de ce premier décodeur de voix à la figure 1.1. (inspiration : [Ferretti-Cinare, 1983]). Ce Vocoder (ou vocodeur) permet de réaliser enfin une analyse intéressante et une compression valable de l'information contenue dans la voix pour la transmission. Le signal vocal entrant dans le décodeur de mélodie est décomposé en dix tensions correspondant à des tranches de 300 Hz; ces tensions peuvent être combinées pour composer un spectre vocal (0 à 3000 Hz). Toutefois, avec le taux de compression obtenu, la transmission et la reconnaissance de signaux vocaux restent encore médiocres.



Depuis 1939, les différentes révolutions informatiques ont engendré différents types de vocodeurs et de générateurs de sons. Aujourd'hui, chaque informaticien peut trouver sur le marché le circuit de parole qui lui permettra de transformer sa propre station de travail en un terminal à sortie vocale de qualité.

La synthèse vocale atteindra d'ici peu son véritable âge adulte. Par contre, la reconnaissance vocale ne possède pas le passé de la synthèse vocale. Les premiers travaux scientifiques sur la reconnaissance automatique de la parole remontent à une cinquantaine d'années et ceux-ci n'ont jamais reçu la caution qu'obtinrent les travaux de développement de la synthèse vocale étudiée en parallèle avec les calculateurs que l'on nomme aujourd'hui ordinateurs.

Après avoir succinctement donné les grandes étapes de la technologie vocale, nous allons décrire plus précisément les orientations principales de celle-ci.

1.3 Les grandes orientations

La technologie vocale se décompose en quatre orientations : la synthèse vocale, l'enregistrement et la digitalisation de la voix, la reconnaissance vocale et la reconnaissance du locuteur.

Nous allons décrire ces quatre orientations en précisant quelques caractéristiques principales.

1.3.1 La synthèse vocale

La première orientation de la technologie vocale, la synthèse vocale, permet à une machine de produire ou de reproduire des sons ressemblant à la voix humaine. Ces sons peuvent être agencés pour former les composants essentiels de la voix humaine [Ferretti-Cinare, 1983, pp 15-24]. Selon la technique employée, les synthétiseurs vocaux peuvent faire entendre :

- des combinaisons d'ondes fondamentales et de bruits blancs afin de produire des résonances particulières appelées **formants** ou de générer des compositions de ceux-ci (par exemple : le formant correspondant à 300 Hz combiné avec le formant correspondant à 2400 Hz donne un son ressemblant à la voyelle i, les bruits blancs servant à produire les **consonnantes** (sons émis pour constituer les consonnes);

- des **phonèmes** (unités élémentaires des sons du langage) : variant d'une langue à l'autre, ils permettent de reconstituer tous les mots parlés de cette langue, (par exemple : en français le nombre de ces phonèmes est de 37 et comprend 16 phonèmes pour produire les voyelles, 18 pour les consonnes et 3 pour les semi-voyelles);

- des **diphonèmes** (segments de la parole qui s'étend de la zone stable d'un phonème à la zone stable d'un autre phonème et qui contiennent en leur milieu une zone de transition entre les deux phonèmes successifs) : il s'agit en réalité des juxtapositions possibles de phonèmes dans lesquelles apparaissent les transitions sonores spécifiques à chaque combinaison de phonèmes (par exemple : le diphonème cha peut représenter facilement le terme chat mais le phonème "ch..." strictement juxtaposé à "...a" ne donne pas le mot chat mais plutôt une approximation du type "chh..a"). Selon le Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI) à Orsay, 627 diphonèmes seraient susceptibles de reproduire n'importe quelle phrase française;

- des **logatomes** (mots sans signification formés d'une consonnante ou d'une combinaison répétée de voyelles et de consonnantes) : ces éléments phonétiques sont une particularisation du concept de diphonème (par exemple : les

logatomes suivant "infor", "mathé", "mati", "que", "cien" permettent de produire des mots tels que "informatique" ou "mathématicien"). Cette technique s'oriente vers une digitalisation d'une voix humaine ou plutôt d'une partie des mots de la langue utilisée;

- des mots ou même des phrases entières (quelques applications directes : terminaux à réponse vocale, machines à laver annonçant la fin de leur programme de lavage, tours automatiques à commande numérique annonçant vocalement les mesures de la pièce à usiner, ...).

En fonction des techniques employées, le synthétiseur devra combiner un nombre plus ou moins élevé de composants afin d'arriver à une production valable et compréhensible par l'utilisateur. Plus le composant de base de la synthèse vocale est particulier (par exemple : un mot digitalisé), plus la production de phrases vocales est restreinte, mais plus l'utilisation ou la manipulation des composants est simple.

Aujourd'hui, cette technologie se retrouve partout. Son utilité est cependant souvent mal estimée par les concepteurs. Ils placent la synthèse vocale sans discrimination précise dans des objets quelconques (de l'appareil photographique "too dark - use flash" au distributeur de limonade "insert correct change and make your selection" - "thank you"). [Schneiderman, 1987, p 254]

Michaelis et Wiggins donnent quelques critères d'utilisation de la synthèse vocale ou de la reproduction de voix digitalisées [Michaelis-Wiggins, 1982, pp. 149-178] :

- le message est simple;
- le message est court;
- le message n'a pas de portée temporelle importante;
- le message requiert une réponse immédiate;
- l'input visuel de l'utilisateur est surchargé;
- l'environnement de travail est trop éclairé ou trop sombre, sujet à de grandes vibrations, ou impropre à transmettre des données visuelles;
- l'utilisateur doit pouvoir se mouvoir facilement;

- l'utilisateur peut être soumis à des forces qui rendent la vision inutilisable ou diminuée (par exemple : le nombre élevé de G, ou le manque d'oxygène perturbent plus rapidement le système visuel que le système auditif).

Après avoir donné les principales caractéristiques de la synthèse vocale, nous allons décrire celles de l'enregistrement et de la digitalisation de la voix, deuxième orientation de la technologie vocale.

1.3.2 L'enregistrement et la digitalisation de la voix

Moins attrayante, mais certainement plus rapidement exploitable, est l'utilisation de la digitalisation de la voix humaine à des fins de messagerie ou de communication. Un utilisateur peut donner un message qui sera diffusé à plusieurs personnes, écouter un message qui lui est destiné et y répondre de façon orale. Le message oral peut être transmis à distance via les réseaux de messagerie existants (par exemple : les services de messagerie développés selon la norme X400 du CCITT).

Le processus de conversion de données digitales en signaux analogiques et vice-versa est une technologie aujourd'hui simple, bien maîtrisée et relativement peu coûteuse. Pour plus de détails, le lecteur peut se référer aux travaux de Shannon. [SHANNON-WEAVER, 1964], [DEMORTIER-LUCAS, 1984]

Les applications des deux premières orientations de la technologie vocale présentent une facette : la sortie d'informations (data output). Dans les deux dernières orientations, nous allons décrire l'autre côté de la technologie vocale : l'introduction de données dans un système (data input).

1.3.3 La reconnaissance des sons et la compréhension de la parole

Si la bouche permet d'envoyer des sons chargés d'un certain sens, l'oreille permet de capter ces sons afin d'en tirer quelques conclusions.

La reconnaissance des sons est une technique similaire au travail de l'appareil acoustique humain. Celui-ci permet de reconnaître un bruit quelconque et de lui associer une sémantique simple et particulière. Si ce bruit est suffisamment précis, une machine pourra sans trop se tromper "reconnaître" le son et ainsi réaliser l'action particulière en rapport avec ce bruit.

En appliquant cette idée à la voix humaine, on obtient la reconnaissance vocale qui pourrait permettre à une machine de reconnaître des mots, des phrases spécifiques et d'en déduire des actions à effectuer. Un schéma simplifié illustre ce processus à la figure 1.2.

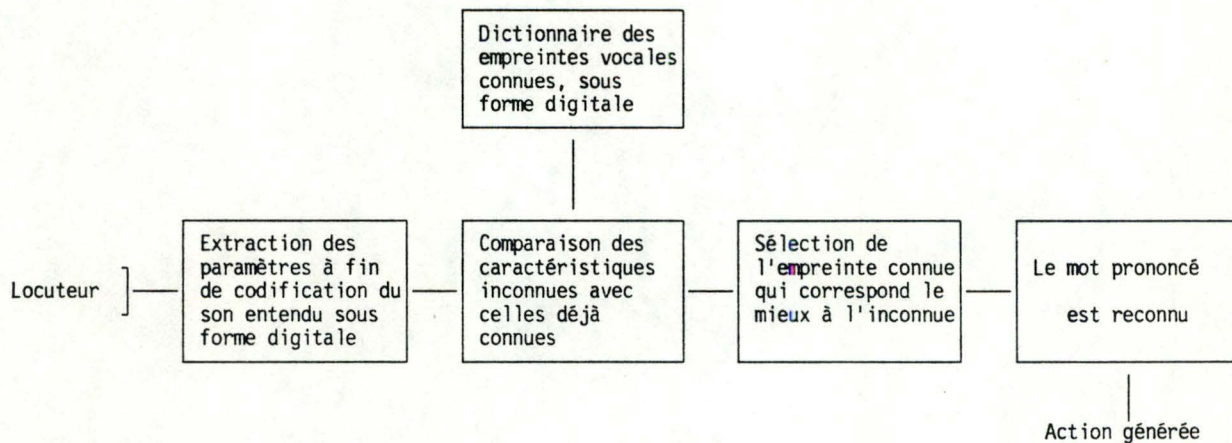


Figure 1.2 : Modèle simplifié de la reconnaissance vocale
2. Sep 1988

Mais la compréhension de phrases prononcées par une personne relève d'un phénomène différent de la reconnaissance vocale. Il s'agit de la reconnaissance sémantique du langage, étape supplémentaire de l'action humaine sur le son entendu, qui associe au son perçu une sémantique plus profonde. Cette étape d'analyse complexe du langage naturel, est employée dans toute conversation humaine. La compréhension qui découle de cette analyse a comme but final de décider d'une action à effectuer, d'une conclusion à tirer. L'analyse pragmatique attribue un sens à la compréhension grâce à diverses considérations liées à la syntaxe employée, aux structures linguistiques, aux indices prosodiques c'est-à-dire à l'intonation utilisée, ...

La séparation de ces deux processus de reconnaissance se retrouve en informatique. D'un côté, pour la reconnaissance vocale, on dispose d'un environnement plus "hardware", comme les "analyseurs de signaux"; de l'autre côté, pour la compréhension, on retrouve des systèmes plus "software" tels les interfaces de langages naturels [Schneiderman, 1987, pp. 57-60, pp. 165-171].

La compréhension de la parole intègre ces deux processus. Elle est souvent confondue avec la reconnaissance vocale bien que cette dernière attache une part moins importante à la compréhension même de la sémantique du signal.

Nous retrouvons cette distinction sur la figure 1.3 décrite par McCauley et parue dans [McCauley, 1984].

COMPREHENSION

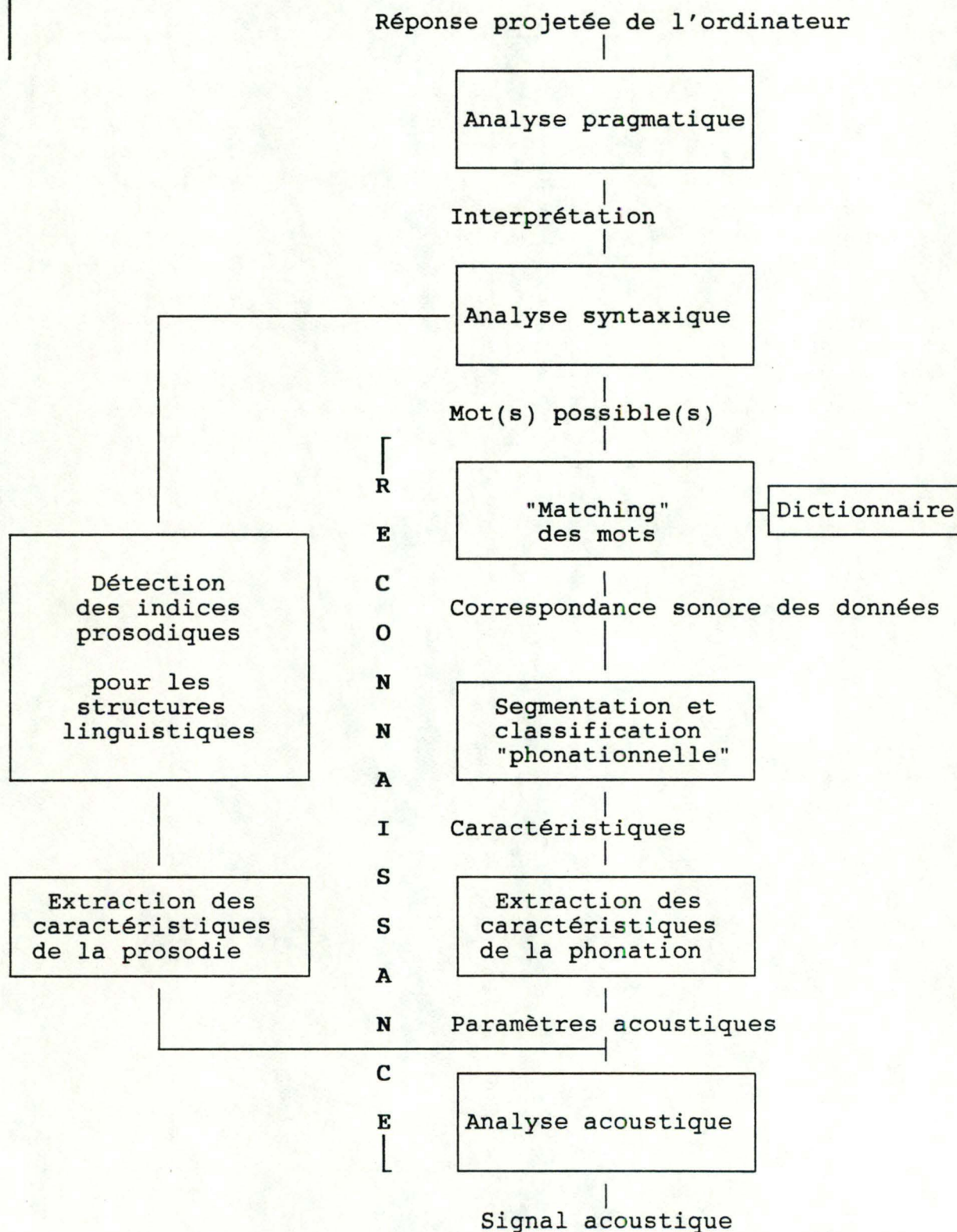


Figure 1.3 : Description d'un système de reconnaissance et de compréhension

Après avoir précisé les principales caractéristiques de la reconnaissance vocale et de la compréhension de la parole, nous allons décrire la dernière facette de la technologie vocale : la reconnaissance du locuteur.

1.3.4 La reconnaissance du locuteur

Au téléphone, il est relativement facile pour chacun de mettre un nom sur une voix bien connue. C'est ce principe, associé à un processus d'analyse du signal, qui permet aux machines de réaliser la reconnaissance du locuteur.

Les systèmes de reconnaissance du locuteur peuvent être considérés comme un cas particulier des systèmes de reconnaissance vocale bien qu'ils soient fondamentalement différents en pratique.

Les applications sont scindées en deux types de système, les systèmes d'identification du locuteur et les systèmes de vérification du locuteur [PHILIP-YOUNG, 1987, pp 8-9].

La technique d'identification permet de reconnaître une personne uniquement à partir d'un échantillon de sa voix et en se basant sur des empreintes vocales déjà enregistrées dans l'ordinateur. Autrement dit le système doit répondre à la question : "Qui est cette personne ?".

Une application de cette technique est l'identification orale de membres d'un groupe d'actions terroristes, criminelles ou d'un groupe de rançonneurs, sur base d'une communication téléphonique par exemple.

La technique de vérification permet à un système de répondre à la question : "Cette personne est-elle bien celle qu'elle prétend être ?". Sur la figure 1.4 (source : [PHILIP-YOUNG, 1987, pp 8-9]), nous pouvons constater que le processus de vérification de l'identité du locuteur s'effectue en deux étapes : la première pendant laquelle le locuteur donne son identité prétendue, la seconde durant laquelle il doit donner un mot de passe ou une phrase-clé sélectionnée selon son identité prétendue pour permettre à l'ordinateur de vérifier si le locuteur est bien la personne qu'il prétend être. La protection de bases de données, l'aide à la réduction de fraudes bancaires par la vérification de l'identité du client, la protection de l'accès réservé à des zones protégées ("restricted area") sont des applications possibles des systèmes de vérification du locuteur.

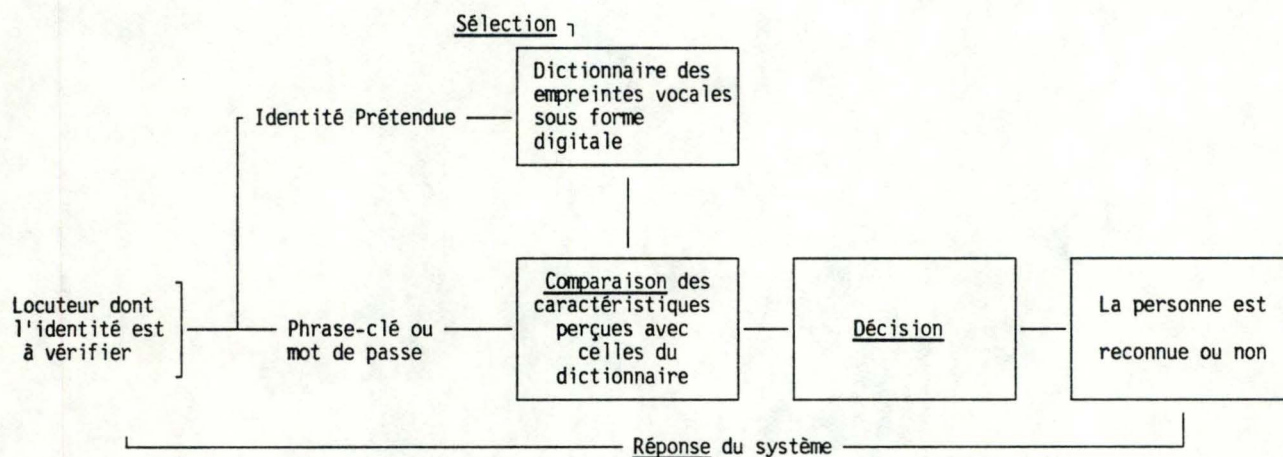


Figure 1.4 : Modèle simplifié de la vérification du locuteur

1.4 Tableau récapitulatif

Nous pouvons retrouver les quatre grandes orientations de la technologie vocale sur la figure 1.5.

La reconnaissance vocale sera détaillée dans le chapitre suivant. Nous y décrirons plus précisément ses caractéristiques ainsi que ses applications.

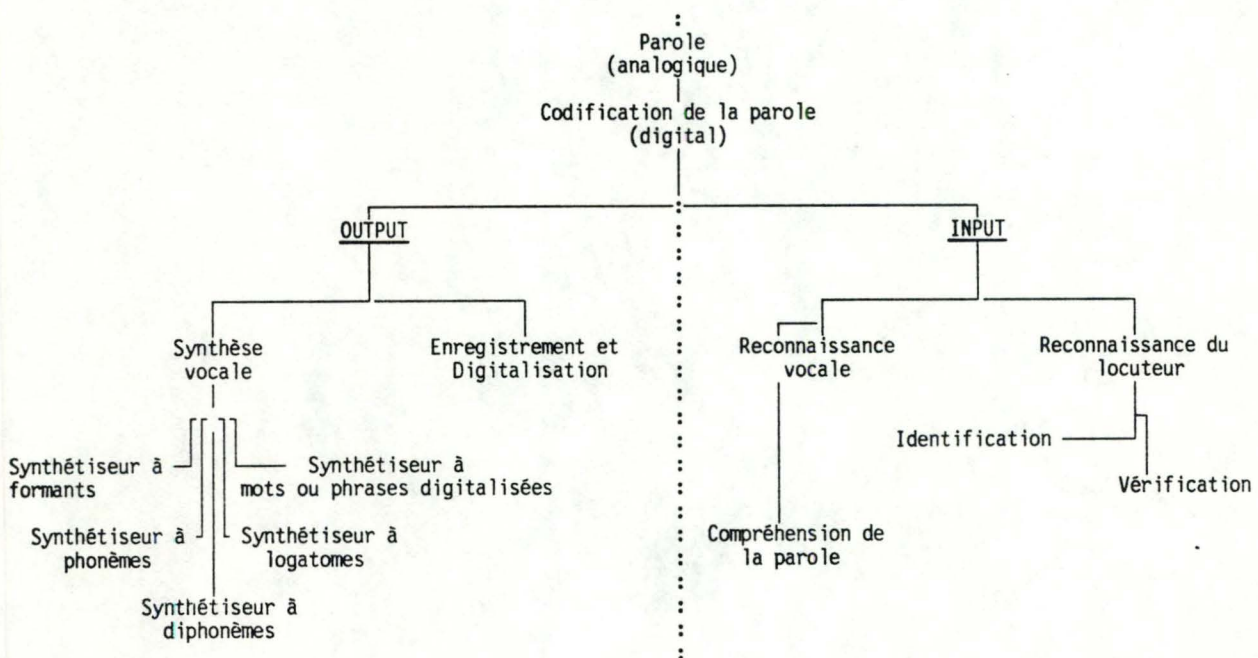


Figure 1.5 : La technologie vocale, ses orientations

C H A P I T R E I I .

La reconnaissance vocale

1. Introduction
2. Les raisons d'utilisation
3. Les grandes étapes
4. La description de la reconnaissance vocale
 - Les caractéristiques
 1. - liées à la parole
 2. - liées à la reconnaissance des formes.
 3. Les méthodes d'analyse et de recherche
 4. Typologie des systèmes de reconnaissance vocale
 5. Les qualités de l'entrée vocale
5. L'état, l'avancée de la technologie (*State of the art*)
 1. Applications du secteur industriel et commercial
 2. Applications télématiques et en télécommunications
 3. Applications administratives et bureautiques
 4. Applications bio-médicales et en ingénierie médicale
 5. Applications dans l'enseignement et en informatique
 6. Applications dans les transports
 7. Applications "grand public"
 8. Applications militaires
 9. Applications diverses
6. Le marché, l'aspect commercial
7. En résumé

II. La reconnaissance vocale

2.1 Introduction

"La roue est une extension des pieds, le livre est une extension des yeux, les vêtements, une extension de la peau, les circuits électriques une extension du système nerveux central."

[McLehan-Fiore, 1967]

"Plus encore que la synthèse, la reconnaissance vocale, avec la commande des machines et des robots, a toujours fasciné l'Homme..."

Qu'il s'agisse d'informatique répartie, de micro-informatique ou de constitution de banques de données, les développements actuels de l'informatique montrent la nécessité d'étudier, tant sur le plan technologique qu'ergonomique et économique, les outils de la communication homme-machine."

[Ferretti-Cinare, 1983, p 25]

Ce deuxième chapitre apporte quelques éclaircissements plus particuliers à la reconnaissance vocale. Les sections suivantes posent d'abord le problème de la reconnaissance vocale en toute généralité : les raisons de son utilisation, un bref historique et ses caractéristiques particulières. Suit une description de l'avancée de la technologie dans différents domaines, à laquelle succède enfin une estimation du marché de la reconnaissance vocale.

Ce chapitre est une synthèse d'articles, de considérations personnelles développées lors de la confrontation avec les systèmes eux-mêmes, d'une recherche pratique sur le marché actuel de la reconnaissance vocale et des réponses apportées par les constructeurs à ce sujet.

Nous retrouverons la liste des articles ainsi que celle des fabricants contactés, respectivement dans la bibliographie et en annexe A (liste des fabricants).

Le chapitre suivant présente un exposé des tests effectués sur le matériel mis à notre disposition.

2.2 Les raisons d'utilisation

Les progrès remarquables dans les vitesses de calcul des ordinateurs et dans leurs capacités de mémorisation de l'information n'ont pas été suivis parfaitement par les innovations dans le domaine des interfaces d'entrée-sortie.

Les téléimprimeurs à 10 caractères par seconde ont été remplacés par des écrans relativement peu coûteux, à affichage rapide (de l'ordre de milliers de caractères par seconde), à résolution raisonnable (moyenne de 320 x 400 pixels), ...

Par contre le clavier, vieux de plus de cent ans reste encore le moyen principal d'introduction de données. Comme l'explique Ben Shneiderman, [Shneiderman, 1987, pp. 228-266], les changements ont été modestes dans ce domaine et ceci est dû à la mauvaise capacité de l'homme à changer et à oublier d'anciennes stratégies de travail. Néanmoins, l'intérêt grandissant vis-à-vis des facteurs humains au niveau des interfaces de communication a débouché sur des centaines de réalisations attribuant une valeur ajoutée aux systèmes classiques "écran-clavier". Par exemple, nous trouvons aujourd'hui sur le marché, des écrans tactiles, des *lightpens*, des manettes de contrôle, des *trackballs*, des souris, ... Une description précise de toutes ces innovations est donnée dans [SCHNEIDERMAN, 1987].

Parmi les hommes, la parole est le moyen de communication le plus répandu. Dans des conditions habituelles, personne ne communique avec son voisin en utilisant un des moyens

artificiels cités ci-avant. L'introduction vocale de données permettrait au moins d'employer un moyen de communication familier à la plupart des humains. De plus, la voix modifie considérablement la position de l'homme vis-à-vis de la machine, en effet, au lieu d'être tributaire de l'écran ou du clavier, il peut se déplacer et accomplir une seconde tâche en attendant d'être averti d'une action à effectuer sur la première. Taper une commande relativement longue sur un terminal peut se révéler fastidieux alors qu'une commande vocale permettrait d'agir quasi instantanément sur la conduite du processus.

Utiliser la reconnaissance vocale comme panacée universelle n'est pas une solution. Comme le décrivent Spериando et Scapin, le problème est de "*choisir le canal sensori-moteur le plus efficient*", de décider quel système de travail appliquer, de déterminer quand la reconnaissance vocale est plus efficace pour l'utilisateur en termes de vitesse de travail, de fiabilité, de résistance au stress dû à la fatigue, à l'appréhension ou à la perte de motivation. [SPERIANDO & SCAPIN, 1987].

Très peu de recherches ont été effectuées sur les aspects ergonomiques de la reconnaissance vocale. Il est donc nécessaire de se reporter à des études relativement anciennes ou aux travaux qui ont été développés sur les autres moyens de communication homme-machine pour évaluer la question.

Avant d'aborder les caractéristiques propres à la reconnaissance vocale, nous allons développer un historique plus précis de ses grandes étapes afin de montrer combien cette technologie est récente.

2.3 Les grandes étapes

- 1943 : M.I. MJASNIKOV publie une analyse sur la
"reconnaissance objective des sons de parole".
- 1944 : J.J. DREYFUS-GRAF débute ses recherches sur le
sonographe, appareil effectuant une analyse spectrale
du signal du son (énergie et fréquence du signal
vis-à-vis du temps écoulé).
- 1951 : les travaux de R. Jakobson, de C.G.M. Fant et de M.
Halle présentent les bases théoriques du détecteur de
phonèmes de S.P. SMITH [R. Jakobson, C.G.M. Fant & M.
Halle, 1951].
- 1952 : K.H. DAVIS, R. BIDDULPH et S. BALESEK exhibent la
première machine à aborder la reconnaissance vocale de
manière globale : les chiffres "0 à 9" sont reconnus
analogiquement avec un taux de réussite relativement
moyen pour une seule voix [K.H. DAVIS, R. BIDDULPH and
S. BALESEK, 1952]. Ces travaux évolueront vers la
reconnaissance de quelques mots [DUDLEY-BALASHEK,
1958].
- 1959 : J.W. Forgie et C.D. Forgie présentent des résultats
quant à la reconnaissance des sons essentiels composant
les voyelles [J.W. Forgie & C.D. Forgie, 1959].
- 1960 : P.B. DENES et M.V. MATTHEWS proposent de comparer
globalement les représentations temps-fréquence,
numérisées et normalisées en durée totale. Cette
méthode permet de reconnaître les dix premiers chiffres
avec un taux d'erreur nul pour un seul locuteur et de
6% pour 5 locuteurs ayant participé à un apprentissage
[P.B. DENES et M.V. MATHEWS, 1960].

1961 : H.F. OLSON et H. BELAR mettent au point un processus de reconnaissance de syllabes phonétiques que le locuteur doit articuler séparément. Le processus est comparable à la reconnaissance par mots : en effet, on peut considérer ces syllabes comme des mots courts dont le répertoire est limité (2000 syllabes couvrent 98 % des besoins de la langue anglaise).

J. DREYFUS-GRAF termine la mise au point du "phonétographe" un appareil de mesure analogique composé de 20 filtres passe-bandes et de circuits identificateurs de phonèmes, utilisant des "compresseurs sélectifs" qui augmentent l'émergence de certains phonèmes. Les résultats sont spectaculaires mais le phonétographe fonctionne avec un seul locuteur qui doit adapter sa diction à la machine.

1965-1970 : Les recherches changent d'orientation. Identifier les phonèmes dans un signal est un problème très compliqué. Les chercheurs s'orientent alors vers deux développements distincts de la reconnaissance vocale. Le premier axe est celui de la reconnaissance de mots isolés (par exemple : pour des applications pratiques qui ne nécessitent que des ordres vocaux brefs, sans explication). La seconde orientation s'attache à la compréhension automatique de la parole, une technique qui utiliserait des informations d'ordre linguistique, lexical et syntaxique pour compléter le message vocal reconnu au niveau sonore (pour construire un dictaphone automatique, par exemple).

1971 : Un tournant est pris dans la recherche sur la reconnaissance vocale. Celle-ci reçoit la caution d'un organisme de recherche important des U.S.A., l'*Advanced Research Project Agency* (ARPA), et voit une réalisation pratique enfin commercialisée.

Un appareil autonome reconnaissant de manière fiable 24 mots isolés après 5 cycles d'apprentissage par le même locuteur, le "*Voice Command System*" de J.W. GLENN et M.H. HITCHCOK est disponible sur le marché.

Un mini-calculateur effectue les traitements d'apprentissage et de reconnaissance, l'analyse du message est effectuée par un banc de 16 filtres, chaque mot est représenté par 8 caractéristiques prélevées au moment de la plus grande variation interne du message.

L'*Advanced Research Projects Agency* (ARPA) finance sur cinq ans un projet ambitieux de compréhension automatique de la parole continue (*Speech Understanding Research*, SUR). Les résultats de ce projet ont été publiés dans *the Review of the ARPA Speech Understanding project* édition du *Journal of the Acoustical Society of America*.

Le but du projet est de reconnaître des phrases prononcées continûment pour des vocabulaires de l'ordre de 1000 mots avec un taux de reconnaissance de plus de 90 %. Mais le projet ne fut qu'un demi-succès (temps de réponse trop long, la reconnaissance devant se faire parfois en temps différé !). Il a néanmoins débouché sur la mise au point de projets opérationnels tels HARPY, HEARSAY II de CMU - *Carnegie-Mellon University* - et HWIM du BBN - *Bolt-Beranek-Newman department*.

En général, la plupart des programmes développés dans le projet ARPA-SUR appartiennent à la catégorie des systèmes de communication en langage naturel restreint à une aire d'expertise précise, c'est-à-dire qu'ils n'utilisent qu'un langage limité.

En particulier, HEARSAY II [V.R. Lesser, 1975], [L.D. ERMAN, 1980], réalisation basée sur Hearsay I [D.R. Reddy, L.D. Erman & R.D. Neely, 1973], est élaboré sur un modèle fonctionnel de modules hiérarchisés selon une relation "utilise" (cfr [AVL, 1988]), c'est-à-dire que chaque module utilise les informations d'un niveau inférieur afin de faire progresser la reconnaissance. Les différents niveaux d'abstraction sont : l'analyse acoustique, l'analyse phonétique, l'hypothétiseur de mots, l'analyse syntaxique et l'élaboration de la réponse.

HARPY est, lui, un système dont l'architecture principale repose sur le parcours d'un réseau sémantique regroupant toutes les informations acoustiques, phonétiques, phonologiques et syntaxiques du message à reconnaître [B.T. Lowerre, 1976].

Quant à la structure du système HWIM, s'appuyant sur des renseignements syntaxiques et sémantiques du domaine linguistique choisi, elle se base sur l'analyse d'îlots de vraisemblance (mot facilement reconnaissable qui permet de formuler des hypothèses sur les mots situés de part et d'autre de celui-ci) [J.J. WOLF & W.A. WOODS, 1977].

Les différents systèmes mis au point font encore l'objet de recherches limitées comme le HWIN et HEARSAY [R. BALZER, 1980] [D.L. McCracken, 1981] ou de développement pour des applications industrielles comme HARPY [LOWERRE & REDDY, 1980] et pour des mises en vente de systèmes plus simples (DRAGON) [FUTURECORP, 1987].

A partir de 1974 : en France de nombreux projets voient le jour.

Au Centre de Recherche Informatique de Nancy (CRIN), J.P. HATON développe le projet Myrtille [HATON J.P. & PIERREL J.M., 1976]. Dès 1975, Myrtille I reconnaît en temps réel de petites phrases continues comme celles échangées dans la demande d'une communication à un standard téléphonique [PIERREL, 1975]. En 1982, la politique de Myrtille II repose sur une reconnaissance de langages pseudo-naturels, dont la syntaxe est proche du français parlé. [PIERREL, 82]. Le système pourrait être appliqué à un centre de renseignements météorologiques.

Au Centre national d'étude sur les télécommunications (CNET), à Lannion, G. MERCIER, P. QUINTON et R. VIVES développent le système KEAL, système de reconnaissance de langages artificiels, conçu pour le dialogue avec une machine en vue d'une

tâche précise : la demande de renseignements et l'introduction d'informations, [MERCIER, 78], [MERCIER et al., 1980], [QUINTON, 1982]. Ce système a abouti à l'élaboration de maquettes expérimentales comme celle d'un centre de renseignements téléphoniques. Les mêmes chercheurs présentent également le système Séraphine qui utilise les principes de la programmation dynamique et applique une méthode globale pour reconnaître des séquences de mots (applications envisageables : numérotation vocale, commande de réseaux locaux, ...).

A Toulouse, le CERFIA propose Arial I, système de reconnaissance multilocuteur permettant de traiter des vocabulaires de taille importante ainsi que son successeur Arial II (1981). Tous deux sont basés sur un modèle élaboré de l'oreille humaine.

Au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), J.J. MARIANI et J.S. LIENARD développent le projet Esope [MARIANI, 1982]. Ce projet étudie des systèmes de reconnaissance basé sur l'analyse statistique des occurrences de signal reconnues et sur une composition prédictive de la solution - reconnaissance de mots à partir de phonèmes de base associée à une induction lexicale. A chaque étape de traitement, les hypothèses les plus probables sont toutes retenues pour vérifier le bien fondé des hypothèses proposées dans l'étape ultérieure. Une application envisagée pour ces systèmes est la gestion d'un agenda vocal (prises de rendez-vous, notes dans le carnet d'adresses, ...).

1980 : le Comité National de la Recherche Scientifique (CNRS) décide de constituer le Groupement de Recherches Coordonnées - "Communication parlée" (GRECO com. par.). Ce groupement a pour rôle de proposer des axes de recherches et de coordonner les différentes études en cours ou en projet.

Notons enfin que les années 1980 virent poindre de nombreuses productions commerciales, comme on le verra dans la liste des constructeurs de systèmes de reconnaissance vocale, présentée dans l'annexe B (constructeurs et produits associés). Actuellement, vu l'évolution de la commercialisation des systèmes de reconnaissance vocale, il est de plus en plus difficile de découvrir des articles récents présentant des nouveautés dans ce domaine, excepté les articles scientifiques de pointe.

2.4 Description de la reconnaissance vocale

2.4.1 Des caractéristiques liées à la parole

Les différents projets qui ont été lancés se sont tous vus ralentis par de nombreuses difficultés directement liées au langage parlé humain. Ces difficultés sont principalement dues au peu de renseignements dont nous disposons au sujet des systèmes auditif, perceptif et de compréhension. Le langage parlé humain est très complexe; la parole est très riche. Comme nous allons le spécifier, cette dernière est particulièrement instable, variable dans ses composants et dans sa forme, continue, redondante, libre et fugitive.

La parole est essentiellement caractérisée par :

- son instabilité : le contenu sémantique d'un message transmis oralement entre plusieurs personnes peut perdre tout ou partie de sa signification originelle (par exemple : le jeu du téléphone arabe, où les participants doivent se communiquer secrètement, oralement et rapidement une information, est une illustration ludique de ce principe de conversion sémantique);
- la variabilité de ses composants élémentaires : les phonèmes - sons élémentaires de la parole - sont soumis à des distorsions différentes selon l'environnement. Les voyelles et les consonnes se prononcent différemment d'après le contexte vocalique;
- la variabilité de ses formes vocales : la prononciation peut varier en fonction du sexe, de l'accent et des habitudes phonétiques de locuteurs d'une même région, de l'état émotionnel et de l'intensité de la voix d'un même locuteur;
- sa continuité : la parole est un flot de mots prononcés plus ou moins rapidement sans terminaison apparente des mots;

- la redondance du signal vocal "porteur" : le discours vocal transmettrait en moyenne 50 bits d'information par seconde, si nous considérons la parole sous forme de transmission de 10 phonèmes (codés sur 5 bits) par seconde. Mais, selon Shannon, [SHANNON-WEAVER, 1964], [DEMORTIER-LUCAS, 1984] [PVBA, 1988] la transmission de la parole ne peut se faire correctement qu'avec un débit minimum de 30.000 bits/s. Plus encore, si une communication parfaite, du type "haute fidélité", est désirée, elle nécessitera un débit de 200.000 à 300.000 bits/s;
- la liberté de sa syntaxe : la syntaxe employée pour le langage parlé est nettement moins stricte que celle utilisée pour le langage écrit;
- sa fugitivité : "Verba Volant, Scripta Manent", les paroles s'envolent, les écrits restent. La phonation est un ensemble de techniques essentiellement momentanées.

Ces caractéristiques suscitent des difficultés qui doivent être surmontées avant d'obtenir une reconnaissance valable.

De par la variabilité des signaux à identifier, il est exclu d'envisager une reconnaissance automatique biunivoque basée uniquement sur les formes vocales et sur une bibliothèque "fixe" d'enregistrements vocaux.

La parole, par opposition à l'écriture, est un phénomène essentiellement continu. De ce fait, il est très difficile d'isoler le début ou la fin des mots, aucun séparateur précis n'existe. Dans le langage écrit les signes de ponctuation, les espaces blancs entre les mots permettent de déterminer de façon précise la fin d'une phrase, d'un mot. Ces séparateurs aident les analyseurs syntaxiques. Mais, inexistants dans le discours oral, ils n'apportent aucune base pratique pour les techniques de segmentation du signal de parole.

La redondance introduit la nécessité d'une paramétrisation de la voix c'est-à-dire de l'extraction du signal vocal d'un nombre restreint de paramètres plausibles, en supprimant l'information redondante.

Pour être réellement utilisables en pratique, les programmes de reconnaissance de la parole doivent être suffisamment évolués pour tenir compte de la syntaxe libre du langage parlé.

La fugitivité de la parole vient renforcer le caractère intemporel et de courte durée du phénomène phonatoire. L'utilisation de la reconnaissance vocale ne conviendra pas à toute application se basant sur une interface nécessitant des données concrètes.

Comme le résume Ferreti et Cinare, [FERRETI-CINARE, 1983], "*la parole sert à véhiculer des idées.*" La voix n'est que "*le signal vocal, le support de la communication parlée...*" Celui-ci (le signal vocal) "*résulte d'un encodage partant d'une idée germant dans le cerveau du locuteur, pour aboutir aux commandes motrices destinées à son appareil phonatoire. A l'aide de son système auditif et de son cerveau, l'auditeur effectue à son tour le décodage inverse.*"

Afin d'effectuer ce décodage inverse, les chercheurs ont proposé de nombreux modèles scientifiques. Les théories développées pour la reconnaissance des formes sous-tendent les systèmes de reconnaissance vocale, comme nous allons le voir dans la section suivante.

2.4.2 Des caractéristiques liées à la reconnaissance des formes

La reconnaissance de la parole s'inscrit dans le cadre plus général de la reconnaissance de formes où l'on entend par "forme" "la description d'un objet choisi dans un ensemble d'objets classés par famille".

En se donnant un critère particulier de classification, il est alors possible de décider si tel ou tel objet appartient ou non à cette classe. Ce travail est relativement aisé pour le cerveau et est assez facile à effectuer mathématiquement, par exemple, en comptant les points anguleux de chaque contour de la forme à reconnaître.

Mais le véritable problème réside dans la difficulté de classer des objets dont le contour est plus ou moins flou ou déformé vis-à-vis de l'image de départ. On définit alors une distance qui quantifie le degré de différence entre l'objet à classer et le ou les types d'objets que l'on souhaite lui attribuer.

Dans le cas de la reconnaissance vocale, on assimile l'objet à classer au signal vocal à reconnaître (le "mot"), la famille d'objets à l'ensemble des "mots déjà connus" et la distance au pourcentage de confiance que l'on accorde au "mot" reconnu.

En pratique, dans des applications de la théorie de la reconnaissance de formes, telle l'analyse automatique de photographies, la reconnaissance des écritures manuscrites ou de la parole, une autre difficulté surgit. Les "images" à analyser sont "mélangées", l'information à décoder est redondante.

Un système de reconnaissance doit être utilisé pour arriver à ces fins.

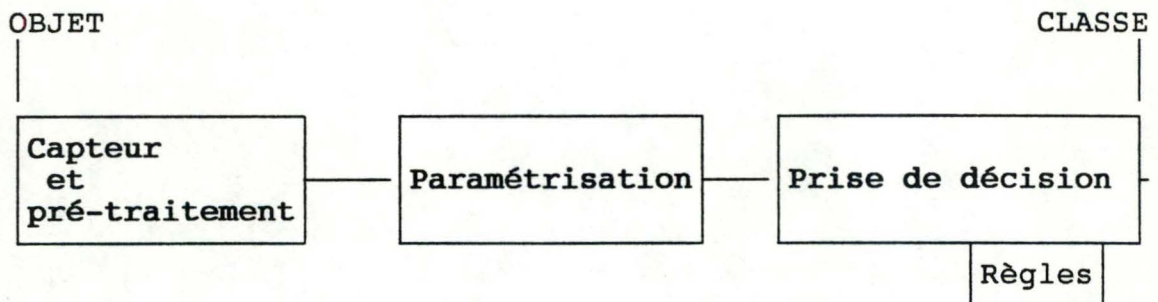


Figure 2.1 : Représentation des sous-systèmes d'un système de reconnaissance de formes

C'est ainsi que tout système de reconnaissance est constitué de trois sous-systèmes, comme le montre la figure 2.1.

Un capteur fournit les "formes" à reconnaître. Un organe de traitement effectue des opérations de lissage, de normalisation, d'extraction de paramètres. Puis, un dispositif décisionnel, sur base de règles, d'images pré-enregistrées détermine la classe probable d'appartenance de l'objet considéré.

Au niveau de la reconnaissance vocale, le **pré-traitement** du signal, **capté** le plus souvent par un microphone, est effectué par un analyseur spectral, un vocodeur ou un appareil de même fonction. Afin de contrecarrer la variabilité des formes vocales, la plupart des systèmes effectuent encore plusieurs opérations, dans l'étape de **paramétrisation**, en vue d'extraire les paramètres réellement utiles. Ces opérations sont entre autres :

- le filtrage du bruit pour éliminer les fréquences hors de la gamme utile, l'élimination du bruit de fond;

- la normalisation de l'amplitude et de la durée des signaux acoustiques;
- l'application de fonctions d'autocorrélation pour éliminer les mesures invariantes par translation;
- la réduction des paramètres à des valeurs "critiques" afin de faciliter un traitement numérique ultérieur;
- l'extraction de caractéristiques particulières : certains phonèmes, certains formants sont faciles à isoler par la présence ou l'absence de certains pics de résonance;
- ...

Pour le processus décisionnel en lui-même, les critères de classification des formes et les règles de parenté des sons à classer sont définis en mémoire.

Les caractéristiques développées précédemment définissent en toute généralité les systèmes de reconnaissance vocale. Mais ceux-ci se différencient les uns des autres par bien des points, comme nous allons le découvrir dans les sections suivantes.

2.4.3 Les méthodes d'analyse et de recherche

A ce jour, nous ne possédons toujours pas la connaissance exacte de la procédure utilisée par l'homme pour reconnaître la parole.

Nous savons que le cerveau humain n'est pas symétrique mais dans le cas de la reconnaissance des formes, les deux hémisphères travaillent simultanément. L'hémisphère gauche est le siège du traitement analytique, de la pensée abstraite, du contrôle des activités motrices, de la génération de la parole. L'hémisphère droit est spécialisé dans la pensée globale, concrète, intuitive et émotionnelle ainsi que dans les perceptions sensorielles.

Confronté à un flot de paroles à comprendre, le cerveau opère sans doute selon une technique de compréhension à divers niveaux. Si, d'après le contexte du discours entendu, le mot à "reconnaître" et à "comprendre" est prévisible, le cerveau opérerait une comparaison globale avec le ou les quelques mots

sélectionnés. Par contre, si le mot n'est pas prévisible ou n'appartient pas à l'ensemble sélectionné, le cerveau ferait une recherche globale dans la mémoire. [IHM, 1988]

Des méthodes semblables sont utilisées pour la reconnaissance vocale. Nous trouvons principalement la méthode de la recherche analytique et celle de la recherche globale.

- 1) la reconnaissance basée sur la recherche analytique a pour but de détecter les éléments phonétiques du message parlé. Elle tire parti de la structure phonétique des mots. Les différentes étapes sont :
 - a) l'extraction des caractéristiques phonétiques
 - b) la segmentation du signal
 - c) l'identification des segments à des phonèmes
 - d) la reconnaissance des mots par les règles phonologiques

La reconnaissance par phonèmes est très attrayante vu le faible nombre de phonèmes existants - une trentaine pour définir la langue française. Théoriquement, cette méthode permettrait de traiter des vocabulaires quasi-illimités, mais pratiquement, elle se révèle relativement médiocre car détecter un phonème en toute sécurité et sans confusion n'est pas toujours possible. L'analyse par vocoder de l'un des phonèmes ne donnera pas toujours le même résultat quel que soit le locuteur.

- 2) la reconnaissance basée sur la recherche globale considère l'unité logique choisie comme une entité globale, non décomposable (par exemple : le mot). Elle procède par comparaison "globale" des unités du dictionnaire de base et des paramètres du signal à reconnaître. Pour ce faire, différentes méthodes sont proposées comme la programmation dynamique, l'analyse par la synthèse du mot à reconnaître, etc... Les systèmes de reconnaissance globale semblent "plus naturels". Ils sont néanmoins confrontés à un nombre d'objets à reconnaître beaucoup plus important ou à

une cascade d'étapes à franchir. Par exemple, afin de reconnaître une phrase complète continue, plusieurs découpages, de plus en plus fins, doivent être effectués, ceci dans le but d'obtenir une signification pour chacun des mots isolés.

Comme l'explique [FERRETI-CINARE, 1983], chacune de ces deux méthodes d'analyse possède des domaines d'applicabilité de prédilection que nous retrouvons dans les tableaux 2.2 et 2.3.

Paramètres du système	Recherche	
	Globale	Analytique
Exportabilité (langue)	bonne	moyenne
Facilité d'apprentissage (petits voc.) (grands voc.)	bonne moyenne	bonne bonne
Coût	raisonnable	coûteux
Utilisation Multilocuteur	moyenne	bonne

Tableau 2.2 : Récapitulatif des domaines d'applicabilité en reconnaissance vocale des techniques globale et analytique selon les paramètres d'un système de reconnaissance.

Objectif Type d'élocution	Recherche	
	Globale	Analytique
Mots isolés - petit vocabulaire	excellente	inadaptée
Mots isolés - grand vocabulaire	moyenne	bonne
Détection des mots	bonne	bonne
Mots connectés Discours continu	bonne	bonne
Phrases avec contraintes ou Langage structuré artificiel	bonne	efficace
Langage naturel	inefficace	excellente

Tableau 2.3 : Récapitulatif des domaines d'applicabilité en reconnaissance vocale des techniques globale et analytique selon le type d'élocution des systèmes de reconnaissance.

Mais dans le processus de reconnaissance n'interviennent pas que des techniques de compréhension, d'association. Nous trouvons aussi des méthodes d'analyse du signal sonore reçu. De nombreuses techniques analogiques et numériques d'analyse de signaux sonores ont été développées. Parmi celles-ci, nous pouvons différencier les méthodes indirectes d'analyse des méthodes directes.

- 1) les méthodes indirectes travaillent à partir d'un signal électrique donné par un microphone, signal construit comme l'image de la mélodie sonore, (variation de la hauteur du son) ou plus particulièrement dans le cas qui nous préoccupe, de l'intonation de la voix. Ce signal est alors confié à l'analyseur proprement dit pour un traitement d'analyse fréquentielle ou pour un traitement temporel selon la méthode d'analyse appliquée. Chacune des méthodes développées utilise des notions importantes de l'analyse du signal (transformées de Fourier, produit de convolution, coefficient d'autocorrélation), d'analyse numérique, de théorie des systèmes linéaires...
- 2) les méthodes directes se basent sur un autre principe : il s'agit de capter de façon directe les vibrations physiologiques engendrées pour produire un son, c'est-à-dire mesurer l'intensité lumineuse transmise à travers la peau du cou (glottographie photoélectrique), mesurer les variations de l'impédance du tissu musculaire du larynx (glottographie électrique), mesurer le temps de fermeture des cordes vocales (glottographie ultrasonore). Le problème souvent rencontré est la modification ou la perturbation des conditions normales de phonation.

2.4.4 Typologie des systèmes de reconnaissance vocale

Lorsque nous sommes confrontés à un système de reconnaissance vocale en cours de développement ou déjà disponible sur le marché, il est nécessaire d'en connaître les caractéristiques pratiques, essentielles pour différencier tel appareil d'un autre.

Deux caractéristiques principales doivent être déterminées en priorité : l'indépendance ou la dépendance du locuteur et le type d'élocution. La première de ces caractéristiques détermine si oui ou non le système peut travailler sur n'importe quelle voix humaine, c'est-à-dire si le processus de reconnaissance est indépendant (multilocuteur) ou non (monolocuteur) des différentes prosodies ou modulations de la voix humaine (accent régional, tessiture, ...). La seconde précise si le locuteur peut énoncer ses phrases continûment ou si le processus de reconnaissance est discret, ce qui oblige le locuteur à séparer chacun des mots à prononcer par des silences plus ou moins longs. Il faut bien différencier reconnaissance discrète de mots, plus ou moins longs, ressemblant même parfois à une phrase et dont la reconnaissance se fait d'un seul bloc, de la reconnaissance continue qui effectue une analyse préalable pour distinguer chacun des mots.

Dans un premier temps, tous les systèmes de reconnaissance vocale peuvent se caractériser par ces deux critères dominants. Une fois ceux-ci déterminés, il nous reste à donner des renseignements secondaires telle la taille du vocabulaire utilisé.

En fonction des critères dominants, on peut attendre des systèmes certaines caractéristiques que nous présentons maintenant.

Pour un système de reconnaissance :

Monolocuteur

Les meilleurs résultats sont espérés, le problème étant plus simple à résoudre;

Le processus poussé à l'extrême permet la vérification de l'identité du locuteur (reconnaissance du locuteur);

Ce type de système nécessite un apprentissage de la voix du locuteur. Le locuteur doit "habituer", "paramétriser" la machine à sa voix. Cette phase d'apprentissage est parfois longue et fastidieuse.

Les applications qui peuvent être soumises à un tel système sont relativement limitées. Ceci est dû entre autres choses au nombre de personnes susceptibles d'être en contact avec l'application. En effet, à chacune des applications développées, il sera souvent nécessaire d'associer un nouvel ensemble d'enregistrements vocaux pour chacun des utilisateurs. Nous nous trouvons confrontés à un problème de ressources à utiliser correctement. La quantité disponible de mémoire des systèmes vocaux n'est jamais illimitée...

Multilocuteur

Les applications auxquelles sera confronté le système pourront être de plus grande dimension.

Mais de moins bons résultats sont attendus, le système devant posséder un référentiel auquel devraient se rapporter toutes les voix. Seul un système utilisant à la fois une référence de base et une modulation de cette référence par les nouvelles voix parviendrait de meilleurs résultats.

Pour un type d'élocution :

Mots isolés / Reconnaissance discrète

Chacun des mots prononcés sera plus aisément reconnu.

Mais ce système témoigne d'un manque de naturel flagrant pour la prononciation d'une phrase complète. Les mots étant reconnus de façon discrète, il sera nécessaire de prononcer une phrase en insérant des silences entre chacun des mots prononcés : "Je ... désire ... parler ... à ... monsieur ... X". Si cette phrase possédait une sémantique propre, par exemple celle d'un mot de passe, elle pourrait être prononcée comme un seul mot : "Je-désire-parler-à-monsieur-X".

Mots "concaténés" / Reconnaissance continue

Le style d'élocution du locuteur est nettement plus naturel. Aucun silence n'est plus nécessaire à la reconnaissance de tous les mots d'une phrase. Enoncer une phrase à vitesse normale suffit à se faire "comprendre" du système.

Mais la concaténation phonique des mots engendre un autre problème, celui des mots parasites c'est-à-dire le risque élevé de prendre un mot pour un autre.

La taille du vocabulaire accepté, caractéristique secondaire, est le nombre maximal d'enregistrements dont le système peut effectuer la reconnaissance. Un vocabulaire réduit permet au système de fonctionner avec une taille mémoire relativement petite, mais les applications sont d'autant plus réduites. Un plus large vocabulaire nécessite une plus importante mémoire de travail. Un microprocesseur performant permettant notamment un accès rapide aux informations est nécessaire si nous désirons un temps de réponse raisonnable (surtout si le système est sensé fonctionner en temps réel).

L'ensemble de ces critères principaux et secondaires reflète les différentes classes de machines qui marquent ou marqueront l'évolution de la reconnaissance vocale. A ce jour, les seuls systèmes qui sont réellement opérationnels sont les systèmes basés sur la communication mot-à-mot. Le plus souvent de type monolocuteur, ils utilisent la technique de reconnaissance de mots isolés. Leur champ d'application regroupe les introductions numériques, les conduites de menus, toutes les tâches qui ne nécessitent pas de véritable dialogue ou qui fonctionnent par question-réponse.

D'autres systèmes, qui commencent à effectuer leur percée dans le monde commercial, sont les systèmes de communication utilisant un langage structuré artificiel. Relativement proche des tâches orientées vers le contrôle de processus, le dialogue entre l'homme et la machine s'installe via un langage comparable aux langages de programmation. Ce langage sert de guide pour la reconnaissance et la compréhension de la parole. Ici, nous trouvons des systèmes répondant à l'ensemble des caractéristiques évoquées précédemment, (par exemple : HEARSAY I, KEAL, ESOPE, ...), les moins évolués restant toujours les systèmes monolocuteurs à reconnaissance discrète.

Les systèmes commerciaux de demain communiqueront avec l'utilisateur via une interface en langage naturel "limité". Nous entendons par "limité", "réduit à une sphère d'expertise précise". Orientés vers l'utilisation par un large public, ces systèmes nécessitent des processus de compréhension et des procédures de dialogue beaucoup plus complexes que les précédents. La majorité des projets définis dans ARPA-SUR devaient aboutir à ce type de système, en particulier HEARSAY II et HWIM.

L'avenir est aux dictaphones automatiques. Aucune restriction n'est imposée aux langages compris par ce type de système, si ce n'est des limites déduites des règles d'élocution. Les routines de reconnaissance n'utiliseront aucune information pragmatique - informations liées au contexte de travail. Le sens des mots ne sera pas choisi en fonction de la tâche à effectuer, mais en fonction de leur sens propre dans le dialogue.

2.4.5 Les qualités de l'entrée vocale

En confrontant l'introduction vocale aux autres moyens de saisies de données, nous pouvons mettre en évidence plusieurs qualités de la reconnaissance vocale.

Damper regroupe et évalue quelques avantages et inconvénients, toujours d'actualité, de la technologie vocale. [DAMPER, 1982] :

Celle-ci serait une solution intéressante dans le domaine des communications à haute vitesse et à faible bande passante (*low band width*), mais les limitations technologiques font que la solution n'est pas encore réalisée en pratique.

Comme nous l'avons déjà décrit, un des avantages majeurs est le fait que si les moyens normaux de dialogue (yeux, mains) sont occupés, la voix permet toujours de commander et de manipuler une machine.

La reconnaissance vocale permet un accès plus aisé par téléphone à une foule de systèmes d'information : *telebanking*, firmes de vente par correspondance, *teleshopping*, gestion et approvisionnement de stock [Krasner, 1986], banques de données diverses, prévision financière [LAZZARO, 1986]. Malheureusement, le réseau téléphonique, en plus de limiter la bande passante du signal, dégrade souvent le signal vocal qui n'est pas toujours correctement perçu.

Ces mêmes remarques peuvent être formulées pour l'usage de la technologie vocale couplée à un accès radio. Avec les émetteurs/récepteurs radio actuels, il est néanmoins possible de libérer complètement l'opérateur de l'ensemble écouteur/micro rattaché à un ordinateur. Un coût additionnel doit alors être compté, mais l'opérateur peut ainsi se mouvoir librement autour de sa machine.

De plus, la technologie vocale est insensible à la quantité de lumière éclairant le plan de travail, ce qui est intéressant dans certaines circonstances comme dans le cadre de travaux photographiques, de manipulations photosensibles.

Nous rencontrons également dans la littérature un avantage vraiment particulier : l'utilisateur n'a pas besoin d'apprendre à taper sur un clavier ni même à lire. Ainsi, un utilisateur totalement novice peut utiliser la technologie vocale. Pourtant, nous en conviendrons, l'utilisateur doit encore apprendre à parler ! A ce jour, l'usage de la reconnaissance vocale n'est pas aussi simple que de parler dans un microphone. Dans les systèmes de reconnaissance vocale les plus communs, l'utilisateur doit exercer sa voix pour se faire comprendre par l'ordinateur, un entraînement parfois astreignant ! La position du locuteur par rapport à la machine est parfois importante. L'ambiance sonore de travail, les perturbations dues aux bruits parasites et aux bruits de fond se répercutent dans certains cas sur les performances du système. Notons quand même que ce dernier défaut est de mieux en mieux corrigé par la technologie actuelle.

Enfin, la parole est un acte naturel pour l'homme. Grâce à la technologie vocale, les hommes et les machines peuvent entretenir un dialogue plus convivial, plus personnel, assez proche pour utiliser le langage naturel. Mais, avec les systèmes actuels, nous sommes loin d'un dialogue "naturel". La compréhension et la génération d'une parole "naturelle" pour des machines "communes" appartiennent au domaine des présomptions pour un futur proche. Mais dès ce moment, ne percevrons-nous pas plus les machines comme un interlocuteur valable plutôt que comme un vulgaire outil et leurs limitations seront certainement moins bien comprises. La reconnaissance vocale possède quelques avantages psychologiques, l'attrait d'une technologie "nouvelle", l'amusement, une certaine motivation de l'utilisateur qui peut contrôler une machine verbalement, comme un esclave. A cette considération, R.I. Damper répond qu'elle est *"d'une grande importance, particulièrement dans les applications ouvertes, mais il y a encore de nombreuses inconnues pour tirer des conclusions irréfutables : comment se sentent les utilisateurs naïfs vis-à-vis du dialogue avec un ordinateur; combien sont intimidés, découragés voire même effrayés par l'utilisation d'une telle technologie ?"* [DAMPER, 1982, p.142]

La technologie vocale entraîne malgré tout quelques désagréments propres à la technique. Souvent, en utilisant un système vocal, nous sommes confrontés à un manque d'information en retour d'une action effectuée verbalement. Si le système donne uniquement ce "feedback" par synthèse vocale, aucune trace persistante de cette information ne pourra être retenue, vu le caractère fugitif de la parole. Un autre problème important est celui de la correction des erreurs. Ces erreurs sont-elles dues à la machine ou à l'opérateur ? Que faire si l'ordinateur apparaît comme sourd ? Dans une telle situation, l'utilisateur peut devenir frustré même irrité, (surtout si la tâche à effectuer est urgente ou importante) bien que la diminution dans la précision de reconnaissance du système ou la surdité de la machine proviennent certainement d'un changement dans l'inflexion de la voix de l'opérateur.

Notons cependant qu'en général, pour de simples tâches, il y a peu de différences significatives dans la rapidité et dans l'exactitude de l'introduction de données via le clavier ou un système de reconnaissance vocale, alors que pour des tâches plus complexes, les performances de l'entrée vocale se détériorent apparemment moins vite que celles de l'entrée par le clavier.

Le cas particulier de la programmation d'un langage informatique (en l'occurrence le SP/4, sous-produit du PL/1) grâce à un éditeur syntaxique utilisant la reconnaissance vocale comme "input device" a été décrit par John Legget et Glen Williams [LEGGETT-WILLIAMS, 1984]. Leurs conclusions sont visiblement optimistes : *"La voix entre raisonnablement en compétition avec l'usage du clavier, bien que l'utilisation de l'entrée vocale soit totalement originale et que les sujets soient expérimentés dans l'usage du clavier"*. Bien que plus lente que l'entrée au clavier, l'introduction vocale produit un taux d'erreur moindre. Le tableau 2.4 résume ces considérations. Les valeurs que nous y trouvons proviennent de l'expérimentation de J. Legget et de G. Williams [LEGGETT-WILLIAMS, 1984]. Le pourcentage de tâche complétée représente la proportion de la tâche réalisée dans le temps imparti pour celle-ci, le pourcentage d'entrée incorrecte, la proportion d'erreurs effectuées lors de l'introduction d'une donnée pour la tâche à accomplir, le pourcentage de commandes erronées, la proportion d'erreurs dans les commandes

nécessaires pour accomplir une action de correction. La mesure d'efficacité de l'édition dénote le nombre de commandes nécessaires pour effectuer une correction valable.

	Editeur-clavier	Editeur vocal
Introduction de données		
% Tâche complétée	70.6	50.7
% Entrée incorrecte	11.0	3.8
Correction de données		
% Tâche complétée	70.3	55.3
% Commandes erronées	2.4	1.5
% Entrée incorrecte	14.3	1.2
Mesure de l'efficacité de l'édition	3.3	3.2

Tableau 2.4 : Pourcentages moyens pour l'introduction et la correction de données via un éditeur "classique" et vocal.

La supériorité de la technologie vocale a été démontrée par Shutoh pour la sélection de choix par menus en particulier dans le domaine de la CAO (Conception Assistée par Ordinateur) [SHUTOH et al, 1984].

Par contre, les mouvements sur écran effectués grâce au clavier sont deux fois plus rapides que ceux ordonnés vocalement [MURRAY et al. 1983].

Suite à ces diverses expériences, nous avons distingué trois types de dialogue quant à l'utilisation de la technologie vocale : les systèmes question-réponse, les menus

de sélection et le dialogue "naturel". Nous détaillons ici les principales qualités de chacun de ces trois types d'entrée vocale.

Le principal avantage du dialogue "questions-réponses" (menu simplifié à une question dont la réponse est supposée connue de l'utilisateur) est la simplicité d'utilisation et de conception. L'ordinateur contrôle totalement le dialogue et ne laisse l'utilisateur poursuivre que si la machine en donne la possibilité. Mais ce style de dialogue homme-machine est parfois gênant. Il ne donne aucun recul sur les différentes alternatives possibles. Il doit être minutieusement conçu. "Les données à introduire doivent être connues de manière très précise lors de la conception du dialogue et leur ordre d'acquisition doit être déterminé à l'avance et imposé une fois pour toutes." [CHANDELON-WARNANT, 1987]

La simplicité et la clarté des menus de sélection, aisément applicables dans de nombreuses situations expliquent leur utilisation courante dans les systèmes actuels. Mais, ici encore, le dialogue reste sous le contrôle de la machine; en outre, il nécessite un support graphique (un écran) afin de présenter les menus. Ce type de dialogue ne convient pas à l'introduction de données qui ne peuvent pas être déterminées à l'avance. Selon J.M. Pierrel, pour limiter le fait que ces systèmes peuvent devenir surélaborés pour un utilisateur non habitué, la solution la plus fréquente à ce jour est de définir des systèmes de sélection par menus cachés qui répondent directement à la requête de l'utilisateur et qui ne montrent les menus sous-jacents que lorsqu'une touche du type "HELP" est pressée. [PIERREL, 1987]

Enfin, d'application et de conception nettement plus difficiles, les systèmes à "dialogue naturel" sont caractérisés par un contrôle du dialogue partagé entre l'homme et la machine. Ils nécessitent des techniques bien plus complexes qui relèvent parfois de l'intelligence artificielle. Toujours selon J.M. Pierrel, ils sont de loin "les systèmes les plus satisfaisants". Il est néanmoins regrettable que ces systèmes n'en soient encore qu'au stade de prototypes.

Toutes ces remarques sur la technologie vocale montrent bien que ce nouveau style de dialogue avec la machine n'est pas seulement un problème à résoudre techniquement. Il est nécessaire, sinon impératif, de considérer attentivement les facteurs humains dans la reconnaissance vocale avant d'avancer une quelconque solution. Pour plus de détails, le lecteur peut se référer à l'ouvrage de B. Schneiderman [SCHNEIDERMAN, 1987] et au mémoire de Muriel Chandelon et Geneviève Warnant sur la modélisation des dialogues [CHANDELON-WARNANT, 1987].

Mais, comme le déplorent J.C. Speriendo et D.L. Scapin, [SPERIANDO-SCAPIN, 1987], *"très peu de travaux ont été effectués sur les aspects ergonomiques de l'emploi de la technologie vocale"*. Vis-à-vis des autres systèmes d'introduction de données, *"il est très difficile de comparer des situations différentes dont les unités de mesure ne sont pas strictement identiques."* De plus, *"l'utilisateur accepte facilement de parler à un ordinateur mais rejette rapidement les contraintes attachées à ce mode de communication"* ainsi que celles, relativement importantes, de la reconnaissance vocale (placement du microphone, entraînement vocal, ...)

2.5 L'état, l'avancée de la technologie (State of the art)

La reconnaissance vocale "discrète" et "continue" pénètre de plus en plus les domaines touchés par l'informatique. *"Il est vrai que de nombreuses applications usant de la reconnaissance vocale sont pleinement réussies si l'une des conditions suivantes est remplie :*

- *les mains de l'utilisateur sont occupées ou inopérantes ;*
 - *une grande mobilité est requise ;*
 - *les yeux de l'utilisateur sont occupés ;*
 - *des conditions rudes (sur un champ de bataille, sous eau) ou des lieux confinés (dans un cockpit d'avion) empêchent l'utilisation d'un clavier."*
- [Schneiderman, 1987, p 250].

Les ergonomes ont défini un autre critère de choix ou de rejet de la technologie vocale pour les domaines d'application. La technologie vocale ne se prête pas correctement à la présentation d'informations permanentes. Celles-ci seront de préférence présentées de façon écrite ou graphique. Par contre les informations temporaires ou à caractère fugitif et qui doivent disparaître après avoir servi conviennent parfaitement à une implémentation vocale.

Comme nous allons le voir, de nombreux secteurs exploitent déjà la technologie vocale, qu'elle soit orientée "synthèse vocale", "reconnaissance vocale" ou "combinaison des deux". Dans la pratique, une réelle évaluation des avantages a été faite et ceux-ci ont été employés au maximum pour diminuer les inconvénients inhérents à la technologie.

Il est bien entendu que l'énumération des différentes applications décrites ci-après n'est pas exhaustif. Chacune d'entre elles est accompagnée d'une abréviation qui spécifie la technologie employée :

(REC)	Système de reconnaissance vocale (input)
(SYN)	Système de synthèse vocale (output)
(REC-SYN)	Système combiné (input-output)

2.5.1 Applications du secteur industriel et commercial

"L'entrée vocale réduite aux systèmes monolocuteurs et à la reconnaissance de mots isolés n'est pas à exclure pour les applications industrielles comme le contrôle de qualité, le tri de colis, le contrôle d'inventaire. La parole discrète est un moyen naturel pour entrer des données ou pour donner des ordres." [BECKMAN, 1982]

En ce qui concerne les applications de la reconnaissance vocale dans le secteur tertiaire, nous les retrouvons au niveau même des machines (machines outils, engins de manutention, système de contrôle de processus de fabrication ou de ligne de production) ou au niveau des appareils de mesure. Combinée à un système de synthèse vocale, des techniciens utilisent la reconnaissance vocale pour le mélange ou le dosage de produits chimiques photo-sensibles qui doivent être manipulés dans une obscurité totale. [VAN PEURSEM, 1984] (REC-SYN)

Quant aux contrôles d'accès à des enceintes de sécurité, la reconnaissance du locuteur est utilisée dans plusieurs entreprises [REHM, 1980]. (REC)

La reconnaissance automatique de la parole se banalise dans le système de CAO (Conception Assistée par Ordinateur). Elle n'exige qu'un faible vocabulaire pour commander différentes tâches (couleurs, formes de dessin, ajustements de position, agrandissements) et elle est suffisamment efficace pour ne pas nécessiter une confirmation systématique du mot reconnu. L'entrée vocale permet au concepteur d'interagir avec le programme de CAO sans toujours faire usage de touches de fonctions spécialisées. [NYE, 1982] (REC)

Chez Technical Instrument Corp., le système de contrôle et d'inspection automatique VIDAS (Voice Interactive Defect Analysis System) est un microscope contrôlé par la voix qui vérifie les circuits intégrés produits au niveau même de la tranche de silicium (wafers). Il permet une augmentation de la productivité et de la satisfaction subjective du travail. [HANSON-WIATROWSKI, 1985] (REC)

Chez Rockwell international, l'inspection de circuits via ce type de microscope accroît de plus de 40 % la vitesse de travail, sans perte de précision ni d'augmentation du taux d'erreur. [MADLIN, 1986] (REC-SYN)

Dans les laboratoires de Biologie, pendant l'étude d'un spécimen, il est plus aisé de continuer un raisonnement sans devoir lever les yeux de la lentille pour écrire des conclusions ou des résultats. Ce qui conduit les chercheurs à utiliser un système qui mélange enregistrement vocal et reconnaissance vocale afin d'éditer automatiquement des rapports (pour des problèmes simples). [PHILIP-YOUNG, 1987] (REC)

La société Westinghouse, *Defense Electronics Plant Departement*, effectue l'inspection de circuits selon une procédure de vérification associant reconnaissance et synthèse vocale. Le système incorpore aussi un lecteur de bar-code et un clavier. Ce dispositif a fait augmenter la productivité et la satisfaction subjective du travail effectué. [MANDEL, 1984]. L'inventaire des entrepôts de la société se fait de la même façon : l'ordinateur demande le numéro du produit et sa quantité en stock. Un manutentionnaire équipé d'un casque et d'un micro FM répond. [MADLIN, 1986] (REC-SYN)

Chez Boeing, les ouvriers manutentionnaires font usage de bras télémanipulateurs commandés vocalement. [SENIA, 1986] (REC)

D'autres systèmes de contrôle d'inventaire peuvent être trouvés dans [Krasner, 1986] et [ROTHBERG, 1980]. (REC-SYN)

Les techniques de maintenance s'orientent aussi vers l'emploi de la technologie de la parole pour libérer les mains de l'opérateur. Afin d'opérer librement des actions de contrôle ou de réparation, un technicien de maintenance porte un casque muni d'un petit téléviseur le dégageant de l'encombrement des manuels. Dans son champ de vision apparaissent, via des commandes vocales, des photos, des séquences de pages décrivant des procédures de maintenance. [KLASS, 1982] [VESTERIG-PROPST, 1982] (REC)

Une autre utilisation du traitement de la parole chez Lockheed Missile System Division, est le contrôle et la conception vocale d'une note descriptive des défaillances pour le matériel électronique à haute sécurité de fonctionnement. [LERMAN, 1980] (REC-SYN)

De même chez Burlington Industries, dans la division de fabrication de verre coulé à Greensbow (Caroline du Nord), l'inspection vocale ainsi que la désignation et la localisation orales des défauts dans le verre a accru la productivité de 25 %. [SEYMOUR, 1986] (REC)

2.5.2 Applications télématiques et en télécommunications

Les télécommunications et la téléphonie se sont également mises à la technologie vocale.

Le téléphone s'équipe de répondeurs et d'enregistreurs automatiques, d'un centre oral de renseignements (dérangements, services RTT ...), d'aide à l'établissement d'une communication, de numérotation et de dispatching, d'un central téléphonique "manuel" automatique. (REC-SYN)

Citons le cas de Bell Telephone Laboratories qui a expérimenté un annuaire téléphonique oral contenant les références de 17 000 abonnés. (SYN)

Le téléphone se transforme en une banque d'informations et de données vocales (météo, horaires et amplitudes des marées, horaires des transports en commun, relevés de comptes en banque); chacune des interrogations à distance pouvant être faite sans terminal visuel, uniquement par voie téléphonique. [LAZZARO, 1986]. (REC-SYN)

Aux Pays-Bas, les PTT ont développé l'unité de réponse vocale PRX 205 de Philips Telecommunicate Industrie BV. Les essais portent sur une unité stockant 80 éléments de parole courante ainsi que les noms de jours et de mois et les nombres de 0 à 20. L'unité a en mémoire quelques règles de syntaxe. La longueur des pauses rend intelligibles les sentences orales composées à partir de ces éléments. Les PTT l'utiliseraient pour la taxation automatique de la communication, la

dénumérotation, le service de réveil et la délivrance de renseignements concernant l'identité et l'adresse d'un abonné dont l'utilisateur connaîtrait le numéro. (SYN)

Aux USA, à l'hôpital de Youngstown dans l'Ohio, un système de renseignements médicaux sur réseau téléphonique gère les tests cliniques requis par les médecins. Une réponse orale présentant les résultats de ces tests leur est donnée par téléphone à l'hôpital, chez eux ou en déplacement. (SYN)

Diverses opérations sont rendues plus sûres et plus performantes grâce à la reconnaissance possible du locuteur, et plus rapides grâce à la commande vocale, par exemple, dans les opérations de télébanking (transfert d'argent, dépôts, retrait; banque "Decatur Federal Bank" de New Orleans) ou de télé-achat. Mais l'impact sera plus important dans ces domaines d'application quand la reconnaissance multilocuteur sera résolue. N'importe qui - accent ou pas, intonation ou pas - devra pouvoir interroger une base de données et se faire comprendre d'un système informatique.

Néanmoins, un système monolocuteur de reconnaissance de la parole peut déjà permettre la réservation de places d'avion, de trains, de théâtre, d'une chambre d'hôtel ou d'une table de restaurant ... en étant correctement utilisé par des professionnels.

Exemple : la reconnaissance vocale est utilisée par les agences de tourisme pour la réservation de places pour les voyages [THOMAS-ROSSON, 1984], [Doddington-Schalk, 1981] (REC-SYN) et pour la consultation automatique des horaires de train [BRUCE-GIBSON, 1984]. (REC)

En France, l'Enserg (de Grenoble) a installé une interrogation orale de base de données informant du trajet des vols, des places libres des vols et du prix du billet.

Il existe également des systèmes automatiques de reconnaissance restreinte basés sur le son constant produit par le téléphone à ton mais ces systèmes sont limités par le clavier du téléphone et obligent l'utilisateur à employer un protocole lourd. Dans les secteurs tertiaire, le TIPS (Telephone Info Processing Service) un téléphone à touches additionné d'un synthétiseur de marque Votrax, permet aux représentants de General Electric d'insérer directement leurs commandes dans les fichiers centraux. Le représentant pianote

sur le clavier numérique du téléphone son numéro de code, sa commande. La machine, sur base des impulsions générées, répond vocalement aux instructions effectuées. Le même principe est d'application chez General Motors et Chrysler, Pfizer, J.C. Penney, La Redoute ... ("REC"-SYN)

IVOIS (Interactive Voice Operated Information System), développé dans l'Illinois, aide un utilisateur novice à consulter l'information nécessaire à la gestion de l'entreprise. *"C'est essentiellement un système d'aide à la décision dans lequel le manager peut interroger une base de données sur ordinateur distant, charger les informations recueillies dans un PC et puis les utiliser avec un des softwares disponibles, tout cela grâce à une simple commande par mot-clé."* [PHILIP-YOUNG, 1987]

2.5.3 Applications administratives et bureautiques

Pour des tâches documentaires et bureautiques, le type de reconnaissance idéal est la reconnaissance du langage naturel.

Pour gérer un secrétariat, il serait merveilleux de posséder un dictaphone automatique. Mais ce dictaphone devrait contenir un système de reconnaissance de la parole continue ayant un minimum de 5 000 mots et cela seulement pour des tâches courantes restreintes. Il faut au minimum 10 000 à 15 000 mots pour un dictaphone automatique et une machine à écrire vocale ! [KURZWEIL, 1986] (REC)

En dépit de ces limitations, la reconnaissance vocale relève peu à peu le défi des tâches de bureau ou de l'administration. [DUSEK-SCHALK-MCMAHAN, 1983], [NAEGLE, 1985], [WALLER, 1984], [WEBER, 1985], [PHILIP-YOUNG, 1987b], [Doddington-Schalk, 1981]. Citons pour exemple l'utilisation de la reconnaissance du locuteur pour assurer la confidentialité des informations ou du mailing vocal à haute sécurité, l'emploi de la reconnaissance vocale pour un système de prise de notes et de mémorandums vocaux, pour la messagerie vocale [DUCHREN, 1985], pour la simplification des fonctions de traitement de texte ou de logiciels "bureautiques" (il suffit de dire "italique" pour que le texte s'écrive en

italique), pour un accès plus aisé à des références bibliographiques.[PHILIP-YOUNG, 1987b] De tels systèmes "intégrés" existent déjà sur le marché. La reconnaissance vocale est employée au sein même de logiciels tels Lotus 123, DBase, Multimate ... pour en améliorer les performances (introduction plus rapide de données). [SEYMOUR, 1986].

Une autre application est le stockage et la reproduction de la parole pour le mailing vocal. Ce système offre aux abonnés un enregistrement des messages vocaux et une réponse aux questions posées par le service. Comme utilisation, nous trouvons la boîte postale téléphonique et la messagerie vocale. IBM, Wang, Delphi Communication Corp. (Exxon), ECS (Electronics Communication System) commercialisent déjà ce type d'équipement. ("REC")

La messagerie vocale serait une solution au jeu du chat et de la souris au téléphone : "Monsieur X est-il là ? Non, essayez plus tard ou à tel numéro. Il est en conférence ...". Beaucoup de temps perdu et d'énergie utilisée en de futilles appels téléphoniques n'aboutissant pas. Les administrateurs perdent 16 % de leur temps au téléphone et les secrétaires 20 %. Deux tiers des appels sont effectués au sein même des compagnies et le pourcentage des appels qui n'aboutissent pas au destinataire au premier essai est non négligeable.

Dans l'administration, la reconnaissance de la signature vocale permet l'autorisation ou l'interdiction d'accès à certaines zones de sécurité.

En criminologie, le système Auros (Automatic Recognition of Speakers) sert à la lutte contre le banditisme. Il est développé au Philips Forschungs Labo de Hambourg. Il se compose d'une base de données comptant 5 000 phrases courtes prononcées par 100 locuteurs ("Mon nom est Némó"). Cette signature vocale est employée aussi pour l'aide policière dans les enquêtes terroristes, dans les affaires d'enlèvement, pour la surveillance du contrôle d'accès, pour l'autorisation d'exploitation et la reconnaissance du locuteur. ("REC")

En conclusion des applications bureautiques, citons Doddington et Schalk :

"You can't automate the office until you automate speech."

[Doddington-Schalk, 1981]

- On ne pourra totalement automatiser la bureautique que si on possède la parole automatique. -

2.5.4 Applications bio-médicales et en ingénierie médicale

La France a mis au point des travaux de commande de robots d'assistance médicale. Ils répondent au signal glottique c'est-à-dire à la vibration des cordes vocales. Un appareil de ce type nommé Glottomat a été employé au Laboratoire Spartacus de l'Inria pour commander un télémanipulateur d'assistance aux tétraplégiques.

Nous pouvons tous constater que le dialogue s'établit déjà difficilement entre personnes dont l'intégrité physique et mentale est complète alors que dire des contacts avec les personnes handicapées.

L'impossibilité d'accès aux grands moyens d'information constitue, pour l'enfant handicapé un retard dans son apprentissage, et pour l'adulte, une déficience dans son développement social. A cela il faut ajouter le handicap lui-même. Pour répondre à ce problème, une série d'outils ont été conçus afin d'aider différents types de personnes handicapées à surmonter leurs déficiences.

Dans cet objectif, le CNET a adapté la reconnaissance de la parole pour aider les handicapés moteurs. Elle leur donne la possibilité de commander leur fauteuil électrique et d'effectuer un nombre limité d'actions sur l'environnement au moyen de la voix (déclenchement de la TV, alarme, téléphone, ...). La reconnaissance d'un petit vocabulaire de mots isolés est suffisant. On peut recourir à la parole naturelle pour les personnes sans trouble de la parole ou se servir d'un vocabulaire artificiel (sons inarticulés) pour les

déficients de la parole. L'outil adapté au traitement de ces problèmes se nomme **Dynamo**. Il reconnaît un vocabulaire d'une centaine de mots prononcés isolément par un locuteur. (REC)

Pour une recherche dans le domaine des applications de la reconnaissance vocale dans l'aide aux handicapés quadriplégiques, le lecteur peut encore se référer à la banque de données [BARD, 1988] et à la Table 2.5 (source : [DAMPER, 1984]).

Une autre application de la reconnaissance vocale à l'aide aux moins valides, chez Boeing cette fois, est l'installation de stations de travail vocal afin d'employer des handicapés pour réaliser des tâches informatiques telles que la programmation et l'analyse. Ce travail s'effectue en couple avec une personne non handicapée. [SENIA, 1986] (REC-SYN)

Le Centre de Recherche Informatique de Nancy (le CRIN), a développé le système Sirène (Système Interactif de Rééducation des Enfants Non-entendants). Son but est d'apporter une aide visuelle pour augmenter la phonation chez un enfant sourd et tenter ainsi de rétablir artificiellement la boucle audio-phonatoire.

Ce système comprend deux parties fondamentales. Nous trouvons un module d'étude des voix pathologiques et un module de rééducation proprement dite. Il est fait appel à des programmes de traitement du signal plus ou moins classiques ainsi qu'à un système de reconnaissance automatique de la parole calculant la différence entre la voix de l'élève et celle du "rééducateur" et cela pour un mot prononcé. Une remarque sur le résultat obtenu par l'élève apparaît. Ce système met en évidence des paramètres difficilement analysables par le jeune (par exemple : la force de la nasalisation). (REC)

L'Institut National des Jeunes Sourds, en collaboration avec le Centre Scientifique d'IBM France, emploie un appareil semblable pour l'assistance à l'éducation de la parole chez des enfants sourds.

Les informations à traiter, c'est-à-dire les voix du professeur et de l'élève, sont captées par un micro, échantillonnées, converties et analysées par un

microprocesseur. Les résultats s'affichent immédiatement sur un écran de télévision connecté au microprocesseur et ils se visualisent par des courbes, dessins animés contrôlés par la voix, jeux ... Ces courbes et jeux aident l'enfant à améliorer la modulation de sa voix ou la prononciation de certains sons. Cet appareil ne demande qu'un clavier à touches permettant de choisir entre différents types d'exercices de la voix. Plusieurs modes de visualisation des résultats sont possibles.

Une thérapie de la parole est explicitée dans [JOST, 1979].

Pour les malentendants déficients de la parole, il existe aussi des machines portables ou adaptées aux communications téléphoniques. Le message introduit chez l'émetteur apparaît sur les écrans de visualisation du récepteur. Citons pour exemple les systèmes Portatel et Portaprint.

Mais les handicaps de l'ouïe ne sont pas les seuls que la technologie vocale peut aider à surmonter. Les déficients visuels en sont un exemple. Des livres parlés existent en cassette audiophonique. La synthèse de la parole devient un moyen intéressant pour la lecture automatique et la première machine à lire à voix haute des caractères imprimés est la machine de Kurzweil. Ceci remonte à 1976. (SYN)

En 1983, le ministère britannique des transports a installé des arrêts d'autobus parlants destinés aux aveugles. Il suffit d'une simple pression sur un bouton et l'heure d'arrivée ou de départ est donnée. A l'intérieur de l'autobus est placé un synthétiseur de voix. (SYN)

Dans un avenir proche, nous pouvons envisager des applications utilisant des signaux physiologiques de trois types (signaux myoélectriques, bio-acoustiques et oraux) afin de commander des systèmes d'aide, des robots d'assistance médicale ou de coopération. Ces systèmes du futur seront commandés par exemple, par la voix, par les ondes alpha générées par le cerveau, par la tension électrique des nerfs moteurs, ...

Applications	Auteurs
Contrôle d'une chaise roulante	[CLARK-ROEMER, 1977] [YOU DIN et al., 1980] [COHEN-GRAUPE, 1980]
Contrôle de l'environnement	[YOU DIN et al., 1980] [COHEN-GRAUPE, 1980] [DAMPER-PURSWANI, 1982]
Traitement de texte	[ODOR-SHARP, 1981] [CREASEY, 1982] [DAMPER-DABBAGH, 1983]
Télémanipulateur	[WICKE et al., 1983]
Contrôle d'un terminal	[GLENN et al., 1976]

Table 2.5 : Références d'applications de la reconnaissance vocale dans l'aide aux handicapés physiques (DAMPER)

2.5.5 Applications dans l'enseignement et en informatique

A l'East Carolina University, USA, un laboratoire de chimie permettant des entrées et sorties vocales a été installé pour les étudiants handicapés. Il consiste en un système portable construit autour de micro-processeurs Z80 et en un système d'acquisition de données analogiques ainsi que diverses interfaces numériques. Le système principal saisit les mesures prises au laboratoire et effectue les calculs

nécessaires au traitement de ces données tandis qu'un système secondaire converse avec l'étudiant. Une des versions de cet outil est destinée aux étudiants aveugles.

Une autre application se situe au niveau de l'apprentissage des langues étrangères. Elle consiste en une reconnaissance de mots isolés et en une vérification de leur bonne prononciation. [SCOTT, 1983]

En ce qui concerne le domaine de l'informatique, John Legget et Glen Williams présentent un système d'édition de programmes par reconnaissance vocale. [LEGGET-WILLIAMS, 1984]

2.5.6 Applications dans les transports

En 1980, la Société Nationale des Chemins de fer Français (SNCF) installe un appareil de composition automatique d'appels téléphoniques sur commande vocale au PC de Reims. Ce dispositif assure la commande centralisée de la voie unique Epernay-Reims et supervise la circulation ferroviaire. Ce système fonctionne grâce à un dispositif monolocuteur de reconnaissance vocale mis au point en collaboration avec la société LCT (Laboratoire Central des Télécommunications). Le but est d'estimer la viabilité du dispositif plongé dans un environnement difficile.

Voici un aperçu des opérations effectuées par le personnel de la SNCF. Chaque régulateur, à sa prise de service, effectue la phase d'apprentissage de l'ordinateur en insérant dans un lecteur les données numériques contenant l'échantillon de sa voix. Dès que les données sont chargées dans l'ordinateur, la phase opérationnelle de reconnaissance commence. Quand le régulateur prononce un nom de gare, l'ordinateur compare la forme vocale émise aux formes-références contenues dans l'ordinateur. Si des formes se rapprochent l'une de l'autre, la voix reconnue et le nom apparaît sur un panneau d'affichage. [FERRETI-CINARE, 1983]

Les transporteurs aériens et terrestres utilisent aussi ce dispositif.

En aéronautique, associée au Limsi, à la société Vecsys et avec le support de la DRET (Direction des Recherches, Etudes et Techniques d'armements, du Ministère français de la Défense), la société Crouzet a développé un processeur de

reconnaissance et de synthèse de la parole couplé avec un simulateur de vol. Ce processeur sert comme organe de commande, d'information et d'alarme dans un environnement et dans des conditions opérationnelles proches de la réalité d'un vol. [ibidem]

Toutefois une utilisation plus intensive dans le secteur de l'aviation est contrainte à une précision plus importante, rendue nécessaire dans un milieu bruyant et saturé en vibrations parasites [COLER, 1977]. Les performances des systèmes de reconnaissance sont telles que bientôt le pilote n'aura plus à bouger les mains dans le cockpit. L'entrée des données (cap, altitude de croisière, etc) s'effectuera par l'interrogation de l'ordinateur de bord. Le pilote connaîtra l'état de santé de l'avion, sans regarder à l'intérieur du cockpit et sans ôter les mains des commandes ou des contrôles importants. [PHILIP-YOUNG, 1987], [FUEGE, 1978], [HARRIS, 1985], [LEA, 1981], [MOUNTFORD, 1980], [WERKOWITZ, 1980] [COTTON, 1983]. Très souvent, le développement de la technologie vocale dans l'aéronautique civile passe par le développement des prototypes militaires [NORTH, 1982], [LANE, 1980], [VAN HEMEL, 1980] dont nous parlerons à la sous-section 2.5.8.

Une application insolite, mais qui appartient aussi au secteur des transports, est celle des ascenseurs à commandes et réponses vocales : vous donnez oralement votre numéro d'étage, l'ascenseur vous y conduit en vous donnant une description de l'étage.

2.5.7 Applications "grand public"

Trois catégories d'application sont définies dans le domaine grand public : les jeux et jouets éducatifs, les équipements domestiques et l'automobile.

En 1978, la reconnaissance de la parole n'en est qu'à ses balbutiements. Mais, déjà, une firme américaine de jouets produisit des camions à 20 dollars obéissant à la voix. Elle reçut 800 000 commandes.

Aujourd'hui, sur le marché, tout le monde peut trouver des interfaces de reconnaissance pour micro-ordinateur. Ces petits systèmes ont un coût raisonnable et conviennent parfaitement à un usage ludique.

Le laboratoire Asulab du groupe Asuag a conçu pour les passionnés de gadget, une montre-bracelet à reconnaissance vocale. Cette montre se compose d'un micro-système permettant la mise à l'heure, le rappel des fonctions et la programmation d'une alarme ainsi que d'un répertoire de 20 mots interchangeables à volonté sans apprentissage particulier.

Outre-Atlantique, un robot domestique a fait son apparition. Il est équipé de radars pour ne pas heurter les murs, d'une carte de reconnaissance vocale pour vous comprendre et d'un synthétiseur de parole pour vous parler. Il passe l'aspirateur, surveille votre appartement et tond même le gazon. Enfin une bonne ménagère !

Par contre, dans le secteur automobile, on n'en est encore qu'aux études et à l'exposition de prototypes. Si la synthèse vocale équipe déjà de nombreux modèles "haut de gamme", il n'en va pas de même pour la reconnaissance vocale. Le Limsi et Renault envisagent des fonctions de commande d'essuie-glaces, de manoeuvre des vitres et de régulation du chauffage. Buick Division (de General Motors) a construit des voitures dans lesquelles l'automobiliste peut commander vocalement l'allumage des phares et des essuie-glaces, la manoeuvre des vitres ... en général des fonctions non vitales. La reconnaissance vocale représente ici une solution à la croissance et à la complexité des tableaux de bord. [MADLIN, 1986] Quant au constructeur Nissan, il permet à l'automobiliste de commander vocalement les rétroviseurs, d'établir sa vitesse plafond ... [YAMAGUCHI, 1983]

2.5.8 Applications militaires

Un programme d'étude, l'AFTI (Advanced Fighter Technology Integration) de l'US Air Force, a étudié le comportement des pilotes des avions de combat F16. Il en est ressorti que la reconnaissance de la parole dans ce milieu, était liée à certaines difficultés supplémentaires dues aux bruits, aux vibrations et au stress du locuteur-pilote. [GODWIN, 1982]

Dans le domaine militaire, la Lockheed Corp. a mis au point le LCC (Lockheed Cybernetics Console), un copilote électronique pour avion de combat. Ce système est appliqué dans des simulateurs de vol. [SENIA, 1986]

Un système d'inspection, créé par General Electric Co. pour l'aviation, utilise le "Dec Talk System" (développé par Digital Equipment corp.) pour les opérations de contrôle sur les réacteurs du bombardier F18 et sur les moteurs de l'hélicoptère Apache [SENIA, 1986]. Des systèmes experts de diagnostic de panne connectés à un système de reconnaissance vocale sont employés pour le matériel aéronautique militaire. [STRIEB-STOKES, 1980]

Enfin, en aéronautique spatiale militaire, les astronautes des navettes de la NASA utilisent la reconnaissance vocale pour diriger les caméras de bord. [SENIA, 1986]

2.5.9 Applications diverses

Dans plusieurs aéroports américains, notamment Denver, le tri des bagages est effectué de manière vocale. Un employé muni d'un micro sans fil, annonce la ville de destination à la machine et l'article sélectionné est envoyé vers la ville désirée. La capacité de traitement atteint 30 000 bagages soit 2 fois plus que le tri réalisé au moyen d'un clavier. La vitesse de traitement et la précision restent identiques quelle que soit l'expérience de l'opérateur. [MADLIN, 1986], [PHILIP-YOUNG, 1987].

Le tri postal est effectué sur le même schéma de travail. [ibidem].

Dans de grandes entreprises, la réservation de locaux est effectuée d'une manière similaire. L'utilisateur donne son nom, le local souhaité, le nombre de personnes à y installer et la destination du local. La machine réserve le local désiré si celui-ci est libre ou, dans le cas contraire, en sélectionne un autre correspondant aux mêmes caractéristiques. [SEYMOUR, 1986]

2.6 Le marché, l'aspect commercial

Réussir à donner une estimation exacte du futur marché de la technologie vocale n'est pas chose aisée.

Voici quelques chiffres glanés dans [FERRETI-CINARE, 83], [SENIA, 86], [SEYMOUR, 86].

Les chiffres avancés sont très disparates. A la lecture de ceux-ci, une seule véritable conclusion s'impose : le marché de la technologie vocale est en expansion !

Les tableaux 2.6 (Le marché mondial (Strategic Inc.)), 2.7 (Le marché américain) et 2.8 (Le marché de la reconnaissance vocale) présentent quelques-uns de ces chiffres.

Les Chiffres

Tableau 2.6 : Le marché mondial (Strategic Inc.)

	1981	1985 évaluation
Valeur totale du marché en \$ de la technologie vocale	38.000.000	645.000.000
Synthèse vocale	23.000.000	495.000.000
Reconnaissance vocale	15.000.000	150.000.000

source : (FERRETI-CINARE, 83), Strategic Inc, (1981)

Tableau 2.7 : Le marché américain

	1982	1984	1987	1995
Valeur totale du marché en \$ de la technologie vocale	26.000.000	105.000.000	781.000.000	1.500.000.000
Synthèse vocale	15.000.000	48.000.000	237.000.000	
Enregistrement et messagerie	6.000.000	32.000.000	274.000.000	
Reconnaissance vocale	5.000.000	25.000.000	270.000.000	

source : (SENIA, 86) et (FERRETI-CINARE, 83), International Ressource Development

Tableau 2.8 : Le marché de la reconnaissance vocale

	1985	1986	1987	1988	1989
Valeur du marché en \$ de la reconnaissance vocale	15.863.000	21.651.000	28.266.000	45.559.000	65.769.000
Nombre d'installation	216.000	422.000	778.000	1.562.000	4.092.000
Shipments	11.400	20.600	35.600	78.400	253.000

source : (SEYMOUR, 86), International Data Corporation

2.7 En résumé

Dans ce chapitre, nous avons présenté l'évolution des techniques de reconnaissance vocale, leur différentes caractéristiques.

Nous avons également réalisé une synthèse des utilisations actuelles de ces technologies dans de nombreux secteurs et des prévisions commerciales qui s'y rapportent.

Nous allons maintenant nous attacher à l'étude approfondie de l'utilisation de deux systèmes particuliers de reconnaissance vocale : le clavier Keytronic KB 5152V et le système VoiceScribe 1000 de Dragon Systems inc.

C H A P I T R E I I I .

L'utilisation et l'expérimentation de deux systèmes de reconnaissance
--

1. Introduction
2. Key Tronic : KB 5152V
 1. Description du matériel
 2. Remarques générales
 3. Comportement à l'exécution
 4. Particularités du système
 5. Conclusions
3. Dragon : VoiceScribe 1000
 1. Description du matériel
 2. Remarques générales
 3. Comportement à l'exécution
 4. Particularités du système
 5. Conclusions
4. En résumé

III. L'utilisation et l'expérimentation

3.1 Introduction

"L'outil apparaît comme un prolongement, un complément de l'être humain, un membre supplémentaire. Bien qu'il ne soit pas moins 'artificiel' que la machine, il bénéficie d'un préjugé favorable. Il paraît simple, soumis immédiatement à la volonté de son utilisateur. Il n'en va pas de même pour la machine. Trois traits me semblent la caractériser et expliquer la méfiance de l'homme à son endroit. Ce sont l'articulation, la répétition et la présence d'une force autonome.

...

L'articulation, cela veut dire que la machine est composée de pièces, qu'elle est capable d'un mouvement interne, à la ressemblance d'un être vivant, cela signifie aussi qu'elle est complexe ...

La répétition, cela implique, d'une part, qu'une machine est en elle-même indéfiniment reproductible, ... et d'autre part, qu'une machine est susceptible de répéter les mêmes effets, les mêmes gestes, à l'infini, sous réserve de son usure.

La présence d'une source de force anime la machine, la dote d'une certaine autonomie..."

[KLEIN, 1974]

Afin de comprendre pleinement ce qu'est exactement la reconnaissance vocale dans la réalité, la meilleure démarche est de s'orienter vers l'utilisation réelle d'un des produits existants sur le marché. Ainsi, confronté directement à la

machine, l'utilisateur peut se faire aisément une opinion personnelle. C'est une des optiques développées dans ce mémoire. La démarche suivie pour y arriver est simple. Possédant un micro ordinateur de la génération des IBM-Personnal Computers (de prix relativement modeste), est-il possible de l'équiper pour la reconnaissance vocale ?

L'annexe B (constructeurs et produits associés) présente une liste de quelques constructeurs qui se sont penchés sur le problème de la reconnaissance vocale.

Parmi les nombreuses cartes ou interfaces de reconnaissance vocale vendues, il est impératif de faire un choix. Pour ce faire, nous devons avant tout cerner les besoins qui poussent une personne à acheter tel type d'équipement et par là déterminer les caractéristiques de "l'interface idéale".

D'une part, un système de reconnaissance "multilocuteur" et "continu" n'est pas toujours nécessaire. Un système de reconnaissance discret et monolocuteur est souvent suffisant. La plupart des applications que nous avons citées précédemment fonctionnent grâce à ce type de système "simple".

D'autre part, les tâches qui seront développées nécessitent-elles un large vocabulaire ? Le nombre de commandes nécessaire au pilotage de l'application est-il important ? Ce besoin caractérisera la taille du vocabulaire du système à choisir. Il n'est pas raisonnable de s'orienter vers le plus petit vocabulaire qui conviendra pour l'application que nous avons en vue et s'il se révèle trop exigu pour un développement ultérieur de celle-ci. De plus, grâce à la technologie électronique actuelle, la taille des vocabulaires des systèmes disponibles est plus que suffisante pour un choix sage.

Une fois les besoins techniques cernés, trouver le produit idéal est encore tributaire du dernier élément qui limitera le choix : le prix.

Mais les systèmes de reconnaissance vocale ne possèdent pas de label de qualité permettant de sélectionner sans erreur une machine de valeur. Ici, la confiance dans la notoriété des constructeurs, leur sérieux sur le marché, l'implication de leur image de marque remplacera avantageusement le tirage au sort parmi les candidats. Il est légitime de préférer un constructeur qui investit dans la recherche, continue à fabriquer de nouveaux systèmes et en publie régulièrement des versions mises-à-jour, à certains fabricants qui n'hésitent pas à essayer de vous vendre leur dernière imprimante ou leur nouveau PC plutôt que de vous donner des informations sur leur interface de reconnaissance vocale.

Parmi les systèmes-candidats restant en lice, nous pouvons dégager trois types de configuration :

- 1° la configuration "externe-clavier" :
une interface externe communiquant avec l'ordinateur hôte via l'entrée-clavier;
- 2° la configuration "interne" :
une carte à insérer dans un des connecteurs d'extension interne (*slot*) dans l'ordinateur hôte (en l'occurrence un IBM-PC);
- 3° la configuration "externe-externe" :
une interface communiquant avec l'ordinateur hôte via une des portes de communication externes (ex : RS-232C).

Deux de ces trois types de configuration ont été utilisées et expérimentées. La première est représentée par un produit de Key Tronic : le clavier KB 5152V, la deuxième par le système VoiceScribe 1000 de Dragon. Le troisième type de configuration n'a pu être expérimenté. Néanmoins, les systèmes de ce type sont assimilables à ceux du deuxième; seule la manière de communiquer avec l'ordinateur-hôte (via un connecteur externe) les différencie. Le lecteur trouvera des représentants de chacun des types dans l'annexe B (Constructeurs et produits associés).

Dans les sous-sections suivantes, nous allons décrire les deux systèmes choisis.

Pour chacun d'entre eux, la sous-section "description du matériel" nous introduit à leurs principales caractéristiques (nom, modèle, fabricant, composition, ...). Puis, la sous-section suivante, "remarques générales", donne quelques remarques préliminaires quant à leur présentation, leur finition, toutes réflexions qu'un utilisateur se fait avant l'utilisation même. Ensuite vient la sous-section "comportement à l'exécution" qui nous explique comment le système se comporte à l'emploi. Enfin, avant de donner les "conclusions" finales sur le matériel testé, la sous-section "particularités du système" présente des remarques quant aux spécificités mêmes du système, des techniques employées ou du logiciel accompagnant.

3.2 Key Tronic : KB 5152V

3.2.1 Description du matériel

Nom : Speech recognition keyboard
Modèle : KB 5152V
Fabricant : Key Tronic
Importateur en Belgique : Manudax, Bld Bockstael, 5, 1020 Bruxelles
Dates de première mise en vente et de révision : 1984 , 1986

Le clavier à entrée vocale (KB 5152 V) de Key Tronic est un représentant du premier type de configuration "externe-clavier".

La société Manudax, importatrice des produits Key Tronic, nous a aimablement prêté le clavier pendant deux mois afin de nous permettre de juger de la qualité du matériel.

Le système complet se compose du clavier, d'un microphone placé sur un serre-tête, d'une disquette comportant la partie logicielle du système et d'un guide d'installation et d'utilisation de 64 pages.

La partie hardware du système de reconnaissance vocale se réduit à un clavier un peu plus lourd qu'un clavier normal et dont la face arrière est munie de connecteurs pour micros et pour interrupteur de micro. Grâce aux logiciels accompagnant le clavier, et une fois l'exécution de ceux-ci lancée via un interpréteur BASIC, un programme permet l'apprentissage de la voix de l'utilisateur par l'ordinateur. Une fois cet apprentissage terminé, l'ordinateur charge, dans la mémoire du clavier, les empreintes vocales nécessaires à l'application, empreintes à partir desquelles s'effectuera l'analyse du signal.

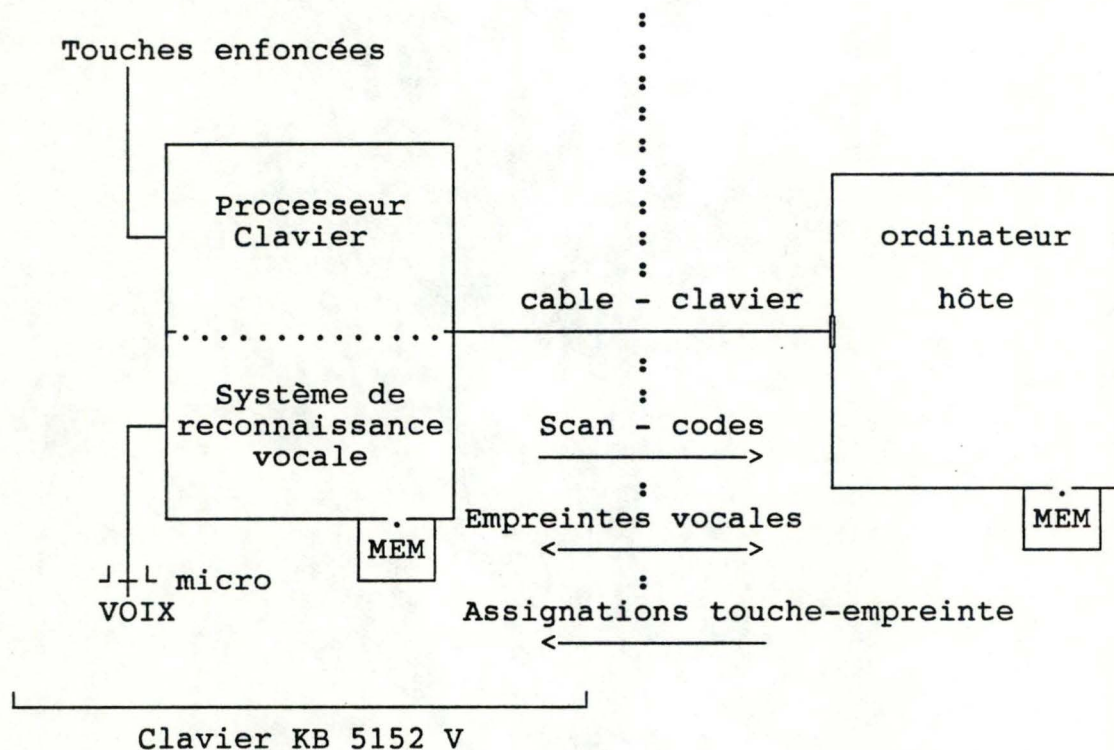


Figure 3.1 : Représentation du système KB 5152 V

Une séquence de touches est assignée par logiciel à chacune des empreintes vocales. Lorsqu'un mot spécifique est reconnu, les touches associées sont communiquées à l'ordinateur via le processeur clavier. Comme indiqué sur la figure 3.1, l'utilisateur peut introduire des données vocalement ou simultanément par la voix et le clavier. Dans ce cas, il faudra alors se méfier des collisions de touches, le processeur clavier envoyant à la fois les "touches-clavier" et les "touches vocales".

Le clavier de type qwerty américain possède, des groupes de touches permettant un contrôle différencié de certaines fonctions :

- Dix touches de fonction (F1-F10)
- Un clavier numérique réunissant les fonctions standards (chiffres, + -, enter et contrôle du curseur)
- Un clavier réservé au contrôle du curseur (copie du clavier numérique mais sans chiffre)
- Un clavier alphanumérique complet sans autre addition de touches (57 touches)
- Des touches isolées pour un usage spécifique :
 - "pause" équivalent à ctrl/Num lock;
 - "cursr pad" commandant l'accès au miniclavier de contrôle du curseur;
 - "voice" muni d'une diode électroluminescente indiquant si le processeur de reconnaissance est en fonctionnement;
 - "reset" qui utilisé avec ctrl, équivaut au ctrl/alt/delete (réinitialisation du système).

Le guide d'utilisation est édité en anglais. Il comporte une description du clavier, du déroulement typique du programme utilitaire ainsi que des instructions de programmation du processeur de reconnaissance vocale.

Le programme utilitaire KTV (Key Tronic Voice Utility Program) permet :

- d'effectuer une démonstration de la reconnaissance vocale en pilotant, par la voix, les mouvements d'un graphique à l'écran;
- d'éditer (créer, modifier) un vocabulaire c'est-à-dire un ensemble de mots à prononcer ainsi que la séquence des touches qui définit une action à réaliser;
- de transférer un vocabulaire du disque vers le clavier
- d'exercer et de mettre à jour la reconnaissance d'un vocabulaire;
- de tester un vocabulaire exercé;
- de charger et sauver un vocabulaire exercé (prêt à l'emploi);

- de modifier les paramètres généraux du processeur de reconnaissance vocale (taux d'acceptation des occurrences, taux de rejet, niveau du bruit ambiant, différence entre un bruit extérieur et un mot prononcé, frontière d'un mot, durée minimale d'un mot, ...)
- de placer le clavier en mode "reconnaissance de la parole" pour que l'utilisateur puisse utiliser le vocabulaire chargé et exercé.

3.2.2 Remarques générales

Dès le premier essai, quelques remarques s'imposent d'emblée avant même de mettre en route le système de reconnaissance vocale.

Au sujet du clavier :

- les touches sont très sensibles;
- le miniclavier de contrôle du curseur n'est qu'une copie du clavier numérique sans les chiffres et n'apporte aucune aide appréciable à l'utilisateur contrairement à la disposition des touches des miniclaviers actuels. Aujourd'hui, les flèches sont regroupées en un T inversé et sont isolées des contrôles "Insert", "Home", "Page Up", "Delete", "End" et "Page Down";
- à l'emploi, la touche "cursr pad", qui valide le clavier curseur et qui bloque le clavier numérique en mode numérique, se révèle plus astreignante qu'utile dans le cadre d'une utilisation du + et du - non numériques exigée dans certaines applications ou programmes développés pour les Personal Computers;

Quant au manuel :

- à la lecture, le manuel semble clair (voir toutefois les réserves qui suivent). L'utilisateur possède une copie de tous les écrans du programme utilitaire ainsi qu'une explication succincte des instructions-machine envoyées au processeur de reconnaissance;
- l'anglais utilisé est un anglais simple;
- des conseils pratiques et efficaces sont donnés afin d'accroître la facilité et les performances d'utilisation de la commande vocale.

3.2.3 Comportement à l'exécution

Si le matériel en lui-même ne présente aucun défaut grave, il n'en va pas de même pour l'environnement d'utilisation du système de reconnaissance vocale.

Nous avons dû développer des trésors d'imagination afin de réussir à faire fonctionner le clavier de reconnaissance. En effet, le manuel n'est pas précis quant aux options de mise en route. Il place implicitement l'utilisateur dans le cadre d'une utilisation d'un IBM PC XT (4.77 MHz) utilisant l'IBM-DOS version 2.

Nous avons néanmoins testé le KBV 5152 sur trois configurations micro-informatiques différentes :

- 1° un clone PC de processeur Phénix capable de fonctionner à 4.77 MHz ou à 8 Mhz et sur lequel était implanté le système d'exploitation IBM-DOS version 3.00;
- 2° un Personal Computer de marque IBM de type XT et muni du MS-dos version 3.20;
- 3° un autre Personal Computer IBM XT, équipé cette fois du système d'exploitation IBM-DOS 2.20.

Dans le premier cas (clone), après de nombreux essais et erreurs suivis de réinitialisations physiques de l'ordinateur, il s'est avéré que le programme ne pouvait s'exécuter que dans un environnement de travail où l'horloge interne tournait à 4,77 MHz et pas à une vitesse plus élevée. Il est à noter qu'une incompatibilité des clones à ce sujet est toujours à envisager.

Dans le deuxième cas (PC XT, MS-dos 3.20), le clavier s'est parfois subitement bloqué : les touches ne répondaient plus mais la commande vocale restait toujours active.

Par contre, dans la configuration-type, c'est-à-dire la troisième (PC XT, IBM-DOS 2.20), aucun problème n'a été rencontré. La reconnaissance vocale s'effectue normalement. Un mot prononcé donne correctement lieu à la génération des touches associées.

Nous aurions bien voulu tester le clavier sur un PC Olivetti mais nous avons constaté une incompatibilité des connecteurs-clavier.

3.2.4 Particularités du système

Le système de reconnaissance vocale proprement dit se compose de trois constituants élémentaires : le processeur de reconnaissance contenu dans le clavier, le programme de gestion du clavier, le vocabulaire de l'application.

Le système est du type "monolocuteur" et n'acceptant qu'un discours "discret".

Le programme utilitaire KTV propose à l'utilisateur huit tâches lui permettant d'utiliser au mieux le clavier de reconnaissance vocale :

- Démonstration
- Editeur (création et édition du vocabulaire)
- Transfert d'un vocabulaire dans le clavier
- Exercice / mise à jour d'un vocabulaire
- Test d'un vocabulaire exercé
- Sauvetage / chargement d'un vocabulaire exercé
- Programmation avancée des différents paramètres de reconnaissance
- Fin du programme et entrée en mode reconnaissance pour une application particulière.

Chacun des points proposés par le programme, à l'exception de la programmation avancée, a été testé. Voici quelques remarques au sujet de l'expérimentation:

- Le chargement et l'exécution du programme doivent se faire via un interpréteur BASIC. Donner à l'utilisateur une version du programme KTV directement exécutable n'aurait-il pas été une meilleure solution ? Evidemment, cette option a le mérite de fournir la source du programme, ce qui permettrait à l'utilisateur averti de programmer plus tard ses propres utilitaires.
- Il est à déplorer que lorsqu'une erreur quelconque se produit (clavier non initialisé en mode vocal, problème d'imprimante, ...), un message est affiché à l'écran mais aucune interprétation de celui-ci n'est donnée explicitement dans le manuel. Ce dernier est apparemment bâti sur le fait que l'utilisateur ne rencontrera jamais que les écrans décrits et aucun autre.
- Le constructeur devrait transformer au plus tôt l'éditeur du vocabulaire. Celui-ci est inutilisable dans son état actuel. Lors des déplacements du curseur, ce dernier reste fixe à l'écran mais "se déplace" néanmoins dans le texte. Il est ainsi presque impossible pour l'utilisateur de savoir précisément où sera affichée la touche frappée.

- Le programme en lui-même ne comporte pas d'autres inconvénients. Le test du vocabulaire exercé est très utile afin de se donner une idée correcte des mots prononcés et de leur proximité vocale les uns par rapport aux autres ou afin d'accroître le taux de réussite dans la reconnaissance d'un mot mal reconnu.

Le troisième et dernier constituant du système est le vocabulaire de l'application que nous désirons conduire vocalement. Ce vocabulaire est l'ensemble de mots, des commandes à prononcer pour diriger l'application choisie ainsi que la définition de la séquence des touches associées pour réaliser une action. Par exemple en prononçant "texte-en-italique", nous aimerions que notre traitement de texte favori (Framework II en l'occurrence) passe en mode d'écriture italique. Il faudra donc associer le mot "texte-en-italique" avec les touches "{ctrl-t}i" qui réalisent cette commande.

La création d'un vocabulaire adapté à un besoin précis est une tâche à effectuer minutieusement. Il est nécessaire de sélectionner tous les termes ou commandes susceptibles d'être utilisés dans le logiciel choisi et de noter quelles sont les séquences de touches à enfoncer pour effectuer ces actions. Ainsi, un vocabulaire de base se constituera peu à peu.

Puis, selon la carte ou l'interface employée, nous devons traduire ce vocabulaire de base en un vocabulaire compréhensible pour le système.

KTV offre à l'utilisateur un choix de code-options afin de distinguer les différentes nuances qui existent entre les commandes à exécuter :

- un vocabulaire peut se subdiviser en sous-ensembles de mots afin de permettre un usage de mêmes mots dans des cadres différents, une reconnaissance plus rapide du mot prononcé (1 terme parmi 10 est plus rapidement et plus précisément reconnu que 1 parmi 100);

- chaque mot est associé à une séquence de touches du clavier à envoyer à l'ordinateur. Pour décrire ces touches, il suffit d'utiliser le caractère lui-même, son code ASCII ou la valeur de son "keyboard Scan-code";
- des mots particuliers peuvent être communs à tous les sous-ensembles de mots. Ce sont des commandes particulières qui sont toujours disponibles comme celles grâce auxquelles l'utilisateur peut demander de l'aide ou clôturer une tâche;
- l'utilisateur peut définir un terme qui permet de placer le processeur de reconnaissance vocale en arrêt momentané. Autrement dit, une fois ce mot prononcé, le micro est coupé pour toute commande si ce n'est un mot particulier qui remettra le système en écoute;
- certains mots peuvent être associés à une notion de répétition c'est-à-dire que certaines commandes peuvent se répéter sans l'intervention du locuteur, et cela jusqu'à ce que le système le réentende à nouveau. Ceci est très utile si nous désirons programmer une commande associée à une touche qui est souvent répétitive, (par exemple, une des flèches utilisées pour se déplacer à l'écran);
- le système permet, par l'utilisation d'un code particulier, que la reconnaissance d'un mot particulier prête à moins de confusion qu'un autre en augmentant le taux minimum de rejet du mot, (par exemple pour reconnaître à coup sûr un mot important).

Le logiciel que nous avons choisi pour développer un vocabulaire pour le KB 5152 V est un logiciel intégré du nom de Framework. Le maximum d'empreintes vocales que l'on peut enregistrer dans le clavier (et donc la taille du vocabulaire) est de 160. Il est à remarquer que ce chiffre est surestimé lorsqu'un vocabulaire constitué de mots longs est employé. En effet, le chiffre de 160 empreintes vocales est lié à la

taille de la mémoire du clavier, dès lors si l'utilisateur utilise de nombreux caractères afin de décrire les mots à prononcer, le nombre total d'empreintes vocales diminuera.

Il faudra bien choisir les commandes à exécuter, Framework donnant à l'utilisateur l'usage de bien plus de 160 commandes.

Nous trouvons en annexe C (Vocabulaire pour KB 5152 V et analyse de proximité) le texte du vocabulaire employé ainsi que les mots qui sont vocalement proches les uns des autres. Ces résultats ont été recueillis directement grâce à la partie "test" du programme KTV.

3.2.5 Conclusions

Nous pouvons qualifier d'honnêtes les performances du système de reconnaissance vocale KB 5152 V. Le matériel est de très bonne facture mais présente les qualités ainsi que les défauts inhérents à un produit ayant peu évolué depuis sa création. Au moment de sa mise en vente, l'environnement logiciel suffisait peut-être à une bonne utilisation du système. Aujourd'hui, face à d'autres logiciels plus conviviaux, plus facilement exploitables, et malgré l'inexistence d'imperfections graves, il paraît totalement obsolète. Cette obsolescence fait ombre aux performances pures de la reconnaissance vocale du KB 5152 V.

Ce système n'est donc pas à conseiller pour une démonstration, fût-elle succincte, des avantages de la reconnaissance vocale.

3.3 Dragon : VoiceScribe 1000

3.3.1 Description du matériel

Nom : VoiceScribe 1000
Version : Release 2.10
Fabricant : Dragon Systems, Inc en collaboration avec The Cherry Corporation
Importateur en Belgique : Multiprox, P.B. 71, Kerkhoflaan, 14-16, 9300 AALST
Dates de première mise en vente et révision successive : 1985, 1986, 1988 (version 3.00)
Article de présentation : [PC-MAGAZINE, 1987]

Les systèmes de reconnaissance vocale de la famille des VoiceScribe sont des représentants du deuxième type de configuration, la configuration "interne". Pour rappel, une telle configuration correspond à une carte à insérer dans un des connecteurs d'extension interne dans l'ordinateur hôte.

Grâce à une contribution personnelle, nous avons acheté le système qui nous paraissait le plus adéquat. VoiceScribe, matériel développé par Dragon Systems, (fabriqué par The Cherry Corporation - Electrical Products Dept -, en collaboration avec Apricot Ltd et IBM), présente le meilleur rapport qualité-prix sur le marché (février 1988). Dragon Systems est une société qui a été fondée en 1982, par les docteurs J.K. Baker et J.M. Baker Elle a pour but de poursuivre les recherches développées dans le projet SUR de l'ARPA des années 70 auquel les deux fondateurs ont collaboré.

Si le lecteur désire plus d'informations sur Dragon Systems, il peut se référer aux bulletins "Futurecorp" édités par Technical Insight, qui décrivent en détail les caractéristiques des entreprises américaines en vogue. [FUTURECORP, 1987]

Le système complet (version 2.10) se compose d'une carte à insérer dans l'un des connecteurs d'extension interne du Personal Computer, d'un micro placé sur un serre-tête, de trois disquettes comportant la partie logicielle du système

ainsi que d'un manuel d'installation (52 pages), d'un guide d'utilisation (172 pages) et d'un manuel d'aide à la conception d'applications (387 pages).

La partie hardware convertit les signaux captés par un micro en données susceptibles d'être traitées par l'ordinateur.

La partie software correspond à un ensemble de logiciels de base, utilitaires et de diagnostic.

Les logiciels de base sont :

Le Driver

Le Driver est le programme le plus important du système de reconnaissance vocale. Celui-ci, résidant en mémoire, capture les données envoyées par la carte centrale, détermine le début des occurrences des mots à reconnaître, les place dans une file d'attente, les traite selon la grammaire chargée en mémoire et leur attribue un numéro d'identification (dans l'étape d'apprentissage) ou leur associe un numéro (dans l'étape de reconnaissance). Ses fonctions sont accessibles via le logiciel DragonKEY ou via la librairie des sous-routines d'interface.

Le compilateur de pseudo-langage VOCL (prononcer "vocal")

Un des logiciels les plus importants pour le concepteur d'application à interaction vocale est un compilateur de vocabulaire associé à un générateur de code grammaire-machine. Il convertit une description textuelle (écrite en pseudo-langage VOCL) du vocabulaire employé et de la grammaire correspondante en une forme compréhensible par le système.

Il est utilisé lors du développement d'une application et pour générer les fichiers de données-machine (*overlays*) pour le programme interpréteur de commande vocale.

Ces *overlays* déterminent quels sont les mots "actifs" ou en droit d'être reconnus et donnent les règles d'évolution du vocabulaire actif en fonction des mots reconnus, des touches frappées ou des commandes générées par l'application.

L'émulateur de clavier "intelligent" DragonKEY

DragonKEY est un programme émulateur de clavier qui permet de contrôler un logiciel ou moyen de la voix. Au lieu de taper une commande au clavier, vous prononcez le mot attaché à cette commande et DragonKEY, grâce à son interface avec le Driver, reconnaîtra le mot prononcé puis effectuera le traitement correspondant dans l'application (envoi de touches, synchronisation avec une autre application ...)

En fait, ce programme peut être considéré comme un logiciel intégrant les fonctions :

- d'interpréteur de commandes vocales définies par l'utilisateur;
- de menu interactif pour l'apprentissage des commandes;
- de test des mots exercés;
- de test de l'environnement sonore de travail;
- d'affichage de l'ensemble des mots actifs;
- d'ajout de nouvelles commandes;
- de contrôle du microphone (On/Off).

Le langage de commande VCOM

VCOM est à comparer au langage de commande DOS. Les commandes de VCOM servent à manipuler DragonKEY. Ces commandes sont envoyées à DragonKEY :

- directement, comme une commande externe du DOS;
- à partir d'un fichier de type "batch";
- en pressant une touche particulière;
- automatiquement, lors du chargement d'une des grammaires;
- directement à partir d'un programme d'application spécialement conçu pour DragonKEY.

Les programmes utilitaires sont :

Le Package DragonLAB

DragonLAB est une station de travail virtuelle destinée à expérimenter la reconnaissance vocale ainsi qu'à développer des applications basées sur la technologie vocale.

Grâce à une "boîte à outil", DragonLAB permet de :

- définir un vocabulaire et une grammaire pour une application;
- sélectionner le meilleur vocabulaire pour une application particulière selon l'environnement et sur base des résultats de tests;
- entraîner des mots d'un vocabulaire particulier et tester la reconnaissance de séquences de mots définies par une grammaire;
- exécuter des tests et récolter automatiquement des statistiques concernant les performances d'un utilisateur;
- simuler une application contrôlée par la voix;
- créer des vocabulaires "adaptables" c'est-à-dire susceptibles de s'adapter aux changements d'intonation d'un locuteur dans le temps ou à la voix de plusieurs locuteurs (simulation d'un système multi-locuteur en établissant la moyenne de tous les signaux reçus).

The Speech Driver Interface Library

La librairie interface du Driver vocal regroupe des fonctions appelables à partir des langages Assembleurs ou C.

Pour les personnes désireuses de contrôler, de manipuler ou de programmer directement le système de reconnaissance vocale, cette bibliothèque de routines est très utile. Elle constitue un ensemble de sous-programmes simplifiant le contrôle direct du contrôleur-décodeur à partir d'applications "vocales" spécialement dessinées dans ce but (par exemple : le programme Aidami-voc, en annexe).

Les logiciels de diagnostic sont :

Diagnose

Le programme Diagnose vérifie les opérations de base de la carte de reconnaissance vocale en fournissant un descriptif détaillé des contrôles à effectuer en cas de mauvais fonctionnement.

Scope

Le programme Scope donne un schéma graphique, semblable à ceux fournis par un oscilloscope, des signaux d'entrée de la carte (en volt ou millivolt). Un exemple de cette représentation graphique est présentée à l'annexe D (Représentation graphique de la voyelle 'a').

3.3.2 Remarques générales

Le sérieux avec lequel cet outil a été étudié transparait dans chacun des composants de VoiceScribe 1000. La carte est réalisée avec des circuits récents. Les manuels sont imprimés clairement et recouvrent apparemment toutes les questions d'un utilisateur potentiel concernant la description du système, la découverte pas à pas de la technologie vocale, la connaissance technique du système. Ces guides ont été écrits en respectant la majeure partie des critères de convivialité décrits par Schneidermann [SCHNEIDERMAN, 1987, chap IX]. Cependant, quelques coquilles d'impressions dans le guide de démarrage et dans l'aide au développement des grammaires sont à relever et la quantité de papier à lire (et à comprendre) pourrait effrayer le novice (611 pages au total dans la version 2.10 !).

3.3.3 Comportement à l'exécution

La mise en route du système nous a positivement surpris. Une fois la carte insérée dans l'ordinateur et les informations de vérification suivies, VoiceScribe 1000 est prêt à fonctionner.

Aucun problème de compatibilité n'a été rencontré; l'ensemble fonctionne même sur les récents modèles 30 d'IBM.

L'environnement de travail est presque sans défaut. La version reçue des programmes se pilote indifféremment par le clavier ou par la voix. Le programme DragonKEY peut donc se commander vocalement. Nous pouvons par exemple, demander vocalement la mise en route du programme d'entraînement d'un nouveau vocabulaire ou le test de la partie du vocabulaire déjà exercé. Cette conduite vocale n'est pas totale : l'ajout

de mots doit s'effectuer via le clavier, l'utilisateur doit taper en toutes lettres l'orthographe du terme supplémentaire ainsi que les touches ou les commandes qui lui sont attachées. Avec la nouvelle version du système de reconnaissance vocale, toutes les actions de manipulation des logiciels pourraient se faire vocalement.

Pour pouvoir manipuler directement le système avec des programmes comme Lotus ou D-base, des grammaires de base réduite à quelques fonctionnalités des logiciels cités sont fournies avec le système. Mais aucune explication n'est donnée sur les grammaires dans le manuel. Seuls les commentaires (bien fournis) à l'intérieur des sources donnent une explication à leur sujet et permettent à un programmeur averti de développer une véritable grammaire pour les programmes susdits. Dans la version 3.00, cette anomalie aurait été corrigée et d'autres grammaires plus larges seraient venues compléter celles déjà existantes.

Le système de reconnaissance ne présente aucun défaut majeur. Les seules imperfections découvertes dans la version 2.10 ont été, paraît-il, corrigées dans la version 3.00.

La version 3.00 n'étant pas en notre possession au moment des essais, nous ne pouvons que faire confiance à d'autres utilisateurs qui nous ont communiqué ces renseignements.

3.3.4 Particularités du système

VoiceScribe 1000 est un système du type "monolocuteur" extensible à un système "multilocuteur". Il n'accepte qu'un discours "discret".

C'est la qualité des logiciels qui donnent sa pleine puissance à VoiceScribe 1000.

La figure 3.2 représente le processus d'entraînement et de reconnaissance du système. Nous constatons que les logiciels y tiennent une part importante.

Les composants essentiels du processus sont décrits ci-dessous (description provenant du manuel de conception d'application chapitre 2 [VS-1000, 1986]). Pour une meilleure compréhension de chaque point, le lecteur peut se reporter au schéma.

1. Le signal sonore du microphone entre dans la partie analogique de la carte de reconnaissance. Cette partie **amplifie** et **filtre** le signal.
2. Le **Processeur** de reconnaissance, également placé sur la carte, **convertit** le signal analogique en une représentation digitale et localise le début et la fin des mots prononcés. Chaque occurrence est mémorisée sous forme d'une **empreinte** vocale du mot prononcé.
3. La carte envoie ces empreintes vers le **driver**, qui réside dans l'ordinateur-hôte. Le driver prend ces empreintes et les place dans une **file d'attente**. Elles sont utilisées pour deux opérations distinctes : l'entraînement et la reconnaissance.

4. Dans la phase d'**entraînement**, l'utilisateur doit prononcer chaque mot un certain nombre de fois. Le driver place en file d'attente une empreinte par prononciation puis construit une empreinte unique, moyenne de toutes les autres : le **modèle**. Les modèles sont mémorisés dans un fichier pour un usage ultérieur.
5. A l'étape de **reconnaissance**, l'utilisateur prononce des mots pour lesquels le système a déjà construit des modèles. L'empreinte vocale du mot perçu est placée dans la file d'attente et, lorsque l'application est prête à recevoir la commande vocale, le driver prend la première occurrence dans la file. La reconnaissance proprement dite a lieu en temps réel (en même temps que l'utilisateur parle) ou une fois plusieurs empreintes mises en attente dans la file. Pour reconnaître un mot, le driver **compare** l'empreinte avec les modèles en mémoire et fournit au programme DragonKEY le **numéro d'identification** du modèle le plus proche de l'empreinte. DragonKEY émulera les touches en conséquence. Ces touches seront directement reçues par l'application qui effectuera l'opération correspondante.
6. Le vocabulaire d'une application peut être grand, puisque limité seulement par la quantité de mémoire de masse disponible.
7. Le **vocabulaire résident** est l'ensemble des modèles chargés en mémoire centrale pour une application et est limité par la taille de la mémoire réservée pour ces modèles.
8. Le **vocabulaire actif** est l'ensemble des modèles à considérer à l'étape de reconnaissance. La taille du vocabulaire actif détermine la vitesse de réponse du système ainsi que la précision de la reconnaissance. Plus grand est le vocabulaire, plus s'accroissent le temps de réponse et le nombre de modèles susceptibles de confusion. La taille du vocabulaire actif est limitée à 1000 modèles.

9. Dans la plupart des applications, il est inutile de posséder à la fois tous les modèles dans le vocabulaire actif. Habituellement, seul un petit nombre de commandes sont utilisées par l'application à un certain niveau de travail. Par exemple, dans une application par menu de sélection, seuls les mots du menu ont besoin d'être "sélectionnables". Une **grammaire**, - appelée "language" par les auteurs du manuel - , définit l'ensemble des modèles actifs pour un état de l'application ainsi que les règles de passage d'un vocabulaire actif à un autre.
10. Le concepteur spécifie un "language" en **pseudo-langage VOCL**. Une fois compilée, cette grammaire est traduite en un format compréhensible par le driver.

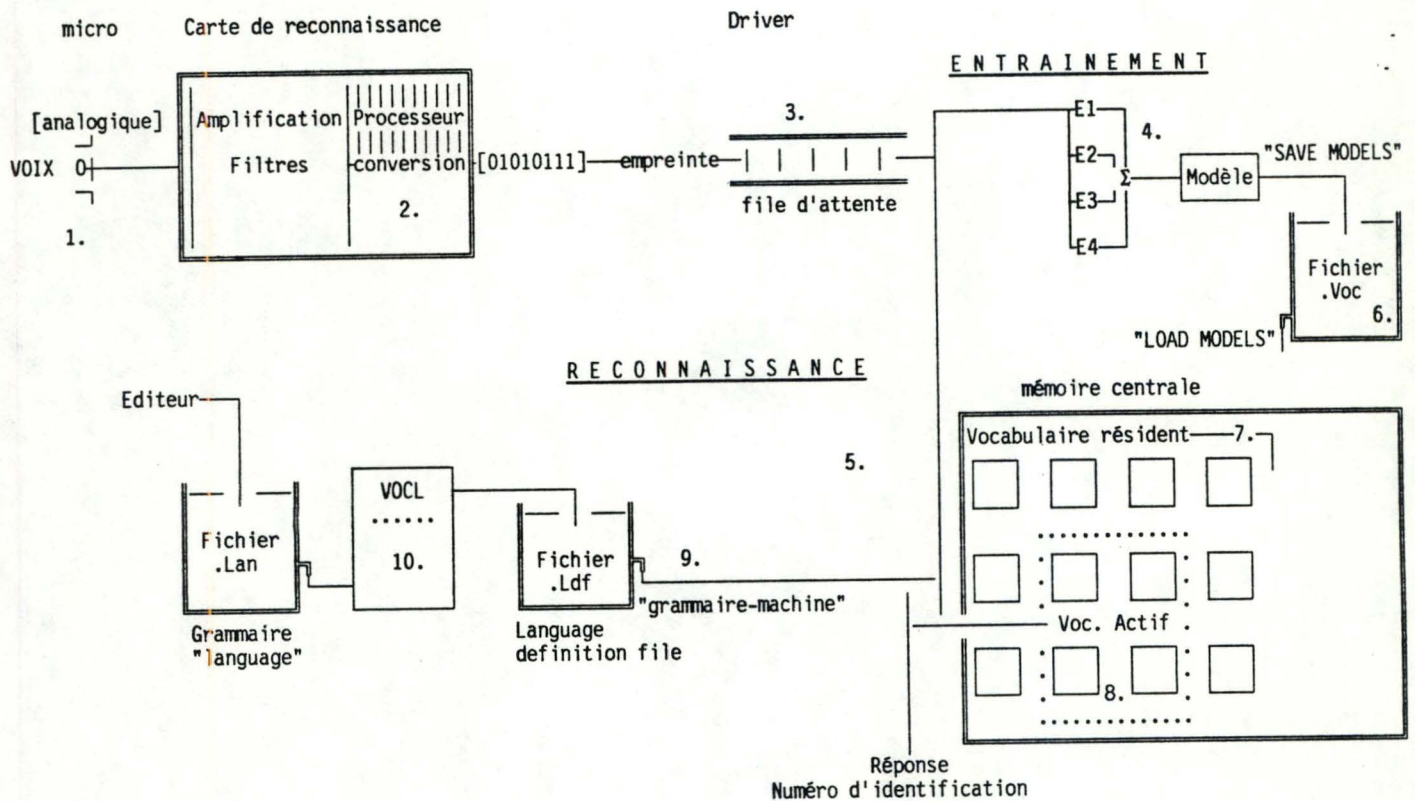


Figure 3.2 : Représentation du système VoiceScribe 1000

Le compilateur VOCL permet à chacun de construire sa propre grammaire pour contrôler l'application de son choix.

La méthode employée pour la spécification des "langages" est une modification de la spécification sous la forme de Backus Naur (Grammaire BNF, "Backus Naur Form"). Les règles grammaticales sont appelées "productions". Elles décrivent la manière de produire des suites de mots, des phrases valides par rapport à la grammaire.

Pour le compilateur du pseudo-langage VOCL, une grammaire est un ensemble de règles qui décrivent les commandes représentant les divers mots à reconnaître et la manière dont ces commandes s'agencent les unes par rapport aux autres. Les seules grammaires acceptées par le compilateur sont les grammaires régulières qui, par définition, génèrent des automates finis [LEROY, 1988]. C'est une représentation de ces automates finis qui est mémorisée et qui sert de contrôleur à la génération des vocabulaires actifs.

Au départ, comme le fait tout automate fini, le système de reconnaissance vocale s'initialise à la production "racine", au symbole distingué, i.e. le vocabulaire actif de départ est composé des symboles terminaux (les mots de la grammaire) directement accessibles en une "transition" (une reconnaissance), une transition correspondant à l'application de règles de production en vue d'accéder à un symbole terminal. Ensuite, le vocabulaire actif évolue selon les symboles terminaux reconnus jusqu'au moment où la décomposition de la production racine en terme de symboles terminaux est achevée. A ce moment, le système se réinitialise en production racine et le processus recommence. Mais des structures de contrôle permettent de changer dynamiquement le déroulement des opérations. Ceci s'effectue grâce à des commandes de la librairie d'interface et n'est réellement applicable que dans les "applications vocales", et non pas dans les utilisations du système VoiceScribe-1000 avec les logiciels du marché (Lotus, Dbase, ...).

En effet, trois types d'applications peuvent utiliser le système de reconnaissance vocale VoiceScribe 1000. Ces trois types d'applications correspondent à trois approches différentes de la conception d'un système de reconnaissance vocale.

L'approche la plus simple (figure 3.3a) est d'utiliser le programme DragonKEY simplement ajouté à un logiciel du marché ou à un logiciel déjà développé. Dans cette méthode d'utilisation, le programme que nous désirons employer n'a pas connaissance de l'existence de l'ensemble vocal. Pour le programme, l'exécution se déroule comme si l'utilisateur tapait directement ses commandes sur le clavier et non pas comme si un autre logiciel lui transmettait les tâches à effectuer. Mais, dans cette approche, il est souvent difficile de garder en synchronisation le vocabulaire actif et les commandes possibles de l'application. En effet, si l'utilisateur effectue, par exemple, un choix de commande via le clavier, la grammaire n'en sera pas avertie et pointera toujours vers le vocabulaire actif précédent qui ne correspond pas à la nouvelle situation. Un autre problème provient directement du logiciel que l'utilisateur désire employer avec la reconnaissance vocale. Si l'enchaînement des commandes du logiciel est à ce point compliqué qu'il nécessite pour l'exprimer, une grammaire récursive, une grammaire ambiguë ou une grammaire dont la décomposition du symbole distingué ne donne jamais que des productions comportant toujours des symboles non terminaux, le système de reconnaissance fonctionnera incorrectement ou pas du tout. En effet, le compilateur VOCL n'accepte ni les grammaires récursives ni les grammaires ambiguës et le programme de contrôle de génération des productions refuse de générer à l'infini des symboles non terminaux.

L'approche de conception la plus intéressante (figure 3.3b) consiste non pas à utiliser les logiciels "clé en main" mis à notre disposition par les concepteurs de Dragon Systems mais à écrire un programme, une "application vocale", contrôlant totalement le driver. Cette approche fait en sorte que le programme profite de la puissance du driver et des facilités d'emploi fournies par la carte de reconnaissance. Cependant notre programme devra inclure toutes les interfaces-utilisateur qui sont offertes par DragonKEY. Par exemple, le programme devra posséder un processus d'entraînement similaire à celui qui est intégré à DragonKEY. Or concevoir une bonne procédure d'entraînement vocal n'est pas une tâche triviale ...

La dernière approche est plus nuancée (figure 3.3c); elle combine les avantages de chacune des deux précédentes.

"L'application vocale" à concevoir contrôle non pas le driver mais DragonKEY par l'intermédiaire d'appels-système au langage de commande VCOM. Comme il le fait avec les programmes "clé en main", DragonKEY transforme chaque commande vocale en simulation de touches clavier. Ainsi, "l'application vocale" peut recevoir des touches comme si elles étaient directement tapées sur le clavier, tout en contrôlant les actions de DragonKEY et en restant de cette façon synchronisée avec les actions de l'utilisateur.

Un exemple de grammaire, l'automate correspondant, ainsi que des valeurs statistiques de reconnaissance se trouvent dans les annexes E (Grammaire pour l'utilisation du Dos avec VoiceScribe-1000) et F (Alphabet et prononciation française).

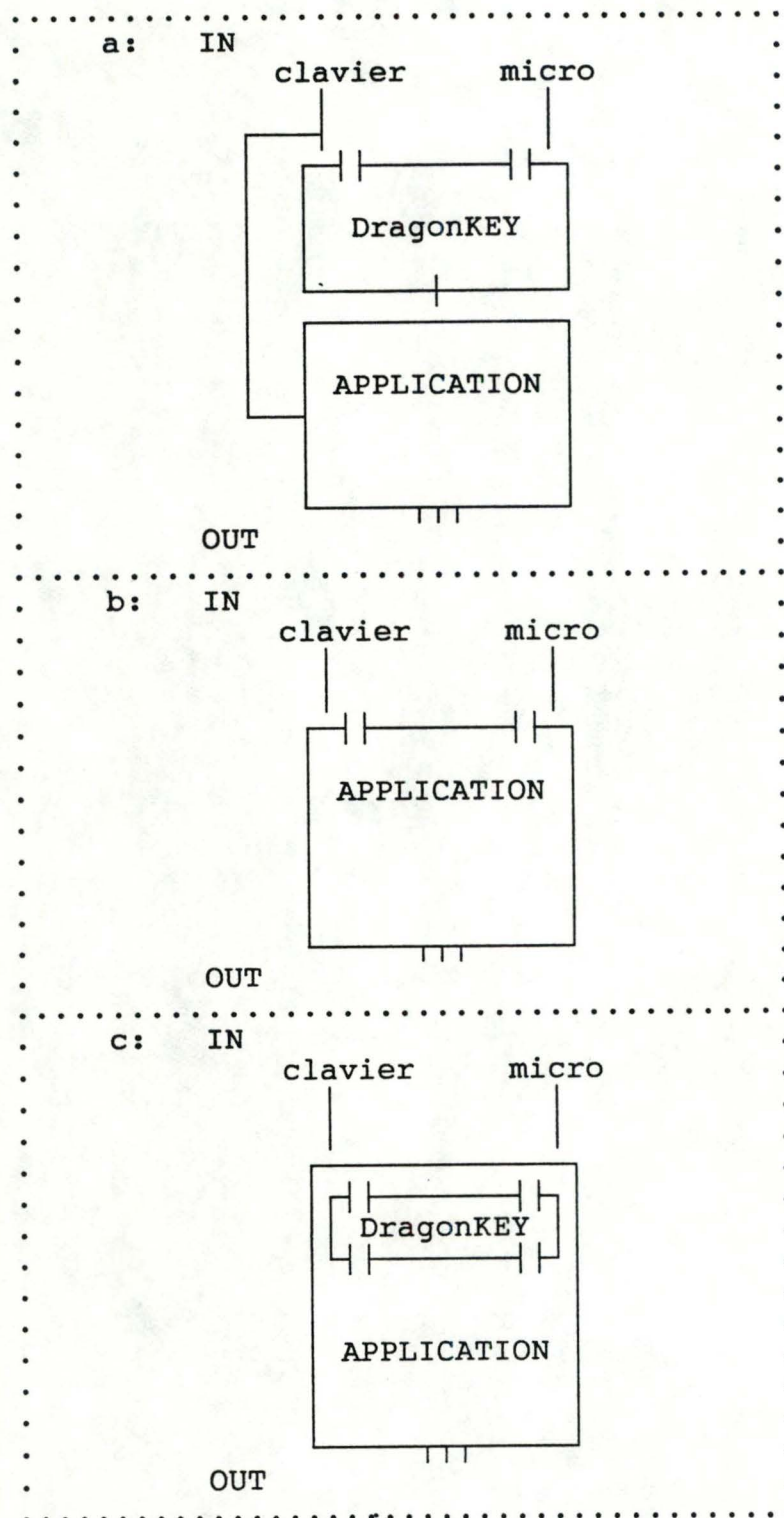


Figure 3.3 : Les trois types d'applications vocales avec VS-1000

3.3.5 Conclusions

Le Système de reconnaissance vocale VoiceScribe 1000 présente de très hautes performances. L'environnement logiciel est particulièrement important. Les programmes sont de très bonne qualité surtout si l'on considère que ces logiciels ne sont pas figés et qu'ils évoluent encore. A l'heure actuelle, si une personne désire "toucher du doigt" ce qu'est la reconnaissance vocale en pratique, VoiceScribe 400 nous apparaît comme une excellente réponse à ses besoins. Cette version est une version tronquée de VoiceScribe 1000. Elle ne peut garder en mémoire que 400 mots à la fois, nombre suffisant pour toute application classique.

En annexe, le lecteur pourra trouver trois applications du système de reconnaissance vocale : une application isolée du type "clé en main" (commande vocale du programme de télécommunication PCTEX et accès au système Videotex), une intégration de plusieurs applications "clé-en-main" associée à un système pseudo multilocuteur (plusieurs personnes peuvent travailler séparément sur un même système de reconnaissance vocale monolocuteur), puis une "application vocale" (Aidami-voc).

3.4 En résumé

Après avoir décrit de manière théorique les systèmes de reconnaissance vocale, nous venons d'effectuer une étude concrète des caractéristiques de deux systèmes actuellement disponibles sur le marché et de leur limites d'utilisation.

Il nous reste encore à particulariser notre recherche en étudiant les possibilités d'adaptation des technologies de reconnaissance vocale à l'aide aux personnes handicapées.

C H A P I T R E I V .

L'aide aux Handicapés

1. Introduction
2. Les handicapés et l'informatique
3. Les handicapés et la reconnaissance vocale
 1. Une mise au point
 2. Des témoignages
4. Le projet Aidami
 1. Les fonctionnalités du système
 2. L'architecture physique et logicielle du système
 3. Remarques et orientations futures

Tableau récapitulatif
5. L'intégration de la voix au projet Aidami
 1. Première étape
 2. Deuxième étape
 3. Troisième étape
 4. Quatrième étape
6. Perspectives d'avenir
7. En résumé

IV. Le cas spécifique de l'aide aux handicapés

4.1 Introduction

"Nous avons découvert que les handicapés, peu importe la gravité de leur incapacité peuvent arriver à un contrôle parfait et global sur un système informatique et l'utiliser normalement.

Le premier pas est de faire face aux besoins spécifiques d'un individu particulier, de cette manière il pourra lui-même faire face à ses propres besoins normaux, identiques à ceux d'autrui."

[HAGEN, 1985]

1989, Europe, 50.000.000 de handicapés de tous horizons...

En 1988, les personnes possédant un handicap ne sont pas encore considérées comme des personnes à part entière. Une fois leur handicap déclaré, elles sont écartées de la vie "normale". Il est vrai que certaines maladies doivent être traitées en milieu hospitalier, que certaines affections doivent s'en remettre à la vigilance de milieux "protégés". Mais faut-il rejeter les moins valides parce qu'ils n'ont pas leur place de "producteur" dans la société ?

Dans les années futures, un plus grand pourcentage de personnes souffrant d'une infirmité resteront des membres productifs de la société. De nombreuses situations incapacitantes à l'heure actuelle ne le seront plus demain. La technologie ne peut guérir une infirmité, mais elle peut l'alléger.

Les Américains l'ont compris depuis longtemps. La section 504 du "Rehabilitation Act" de 1973 impose que les travaux de bureau, les écoles, les bâtiments de la plupart des institutions publiques soient accessibles aux moins valides. Derrière la lettre de cette loi, nous pouvons souligner une tentative de ne pas exclure les handicapés du monde "normal".

De plus, aujourd'hui membre de la commission sur l'examen de la section 508 des "Rehabilitation Act Amendments of 1988" [PUBLIC-LAW-99-506, 1986] au sujet de l'accessibilité des handicapés à tout système informatique, William Roth fait remarquer : *"La première contribution de l'Electronique moderne a été de perpétrer des changements dans la révolution industrielle et de créer une nouvelle définition de l'infirmité. Etre handicapé ne se mesure plus en terme du nombre de kilos qu'une personne peut soulever ou du nombre d'escaliers qu'elle peut gravir mais plutôt en la manière dont elle peut introduire des données au clavier, ou bien encore du fait qu'elle puisse lire des informations sur un moniteur."* [C.H.I., 1988] La seconde contribution est que cette nouvelle technologie est plus facilement modelable aux besoins de l'homme que ne l'étaient les anciennes. Malheureusement ceci ne signifie pas que le changement va avoir lieu de suite. Les mentalités, la politique des pouvoirs publics doivent d'abord changer. De nouveaux marchés commerciaux doivent se créer. Chacun doit ouvrir les yeux face à ce problème.

L'informatique et les informaticiens peuvent apporter leur soutien ...

4.2 Les handicapés et l'informatique

L'informatique a révolutionné le monde des handicapés. Toutes les personnes que nous avons rencontrées dans le cadre de ce travail en conviennent. Aujourd'hui, de nombreux handicapés peuvent affronter plus sereinement les "petits problèmes" de la vie, accéder à une nouvelle autonomie, réapprendre à apprendre ou tout simplement avoir des loisirs.

Que l'informatique soit réduite à un microprocesseur intégré, à un téléphone main-libre ou qu'elle soit directement visible dans les tâches qu'effectue un micro-ordinateur, elle apporte une aide directe ou indirecte à des handicapés de toute nature.

L'utilisation d'un clavier peut être un insurmontable défi pour des personnes handicapées physiquement. Pourtant, 200 maisons "*Cheshire Homes*", principalement établies en Angleterre, utilisent entre autres la micro-informatique comme aide à la réadaptation de grands handicapés physiques [COMPAID, 1988]. Deux centres en Grande Bretagne ont particulièrement développé cette politique d'accès à l'informatique : "*Seven Springs*" près de Tunbridge Wells et "*Heatherly*" près de l'aéroport de Gatwick. Selon Robin Nixon, conseiller pour l'aide informatique aux handicapés dans le centre de "*Seven Springs*", "*le potentiel de la micro-informatique à aider les personnes souffrant d'une infirmité est vraiment plus important que pour les personnes valides, puisqu'elle leur offre un réel moyen de communication.*" Les jeux sur ordinateur sont souvent le point de départ de toute une pseudo-thérapie. Grâce à la ludotique d'abord, les personnes possédant un handicap physique stable - ou évolutif comme au centre "De Bijtjes : Les Petites Abeilles" de Vlezenbeek, Belgique - développent leur dextérité ou leur concentration vis-à-vis d'une tâche à accomplir. Puis, selon leur affinité, leur centre d'intérêt, elles peuvent s'orienter vers des tâches plus "bureautiques". Au centre de "*Seven Springs*", par exemple, les résidents étudient des langues, des matières scientifiques, aident à la comptabilité du centre, à toutes les activités propices à la communication, dactylographient les lettres du secrétariat. Au centre de Kerpape, près de Lorient, ce sont des enfants muets et complètement paralysés qui dialoguent avec le "monde extérieur", jouant entre eux et se disputant même via leurs systèmes de dialogue. Ces enfants vont à l'école, apprennent l'orthographe, les conjugaisons, l'arithmétique, l'histoire, ... Dans un article consacré à l'effort de Kerpape, André Sylvestre, psychologue du centre, s'exprime ainsi : "*Si Jean-Jacques, paralysé complet et muet, parvient sur son micro-ordinateur à un très bon résultat aux tours de Hanoi, cela signifie quelque chose tout de même.*" [TORTEL, 83].

Chacun des handicapés possède sa propre interface de travail adaptée à son infirmité. Celui-ci utilise des interrupteurs placés aux pieds, celui-là dispose d'un joystick oral. Mais les micro-ordinateurs ne sont pas exclusivement réservés à un usage "sérieux", ils permettent à des personnes très gravement atteintes de retrouver la confiance en elles du simple fait de réussir à faire obéir une machine personnellement, sans l'aide d'un tiers. Tous ces gens, souvent astreints à dépendre d'un autre pour toutes sortes de petits travaux domestiques, peuvent enfin accéder à une certaine autonomie. Un robot travaille pour eux. Un système de contrôle de l'environnement leur allume la télévision, règle l'intensité lumineuse de la pièce, leur permet de téléphoner à une autre personne : posséder une partie de l'autonomie du valide, voilà le rêve de nombreuses personnes immobilisées.

Les grands handicapés moteurs ne sont pas les seuls à bénéficier de l'aide de l'informatique. Les personnes atteintes d'une infirmité visuelle peuvent aujourd'hui lire un écran d'ordinateur grâce à un système d'affichage en braille ou grâce à des moniteurs ou des logiciels "agrandisseurs" qui permettent aux mal-voyants d'agrandir les textes présentés à l'écran jusqu'à distinguer correctement ce qui est écrit (Système Braille-Edit-Express décrit par J.J. Lazzaro [LAZZARO, 1986]). Pour les handicapés de la vue, des systèmes permettent aujourd'hui de lire n'importe quel écrit dactylographié. Des appareils de lecture automatique ou de conversion de l'alphabet dactylographique en alphabet braille sont en vente depuis plusieurs années (la Kurzweil Reading Machine, 1979). Pour les personnes ne connaissant pas le braille, des synthétiseurs de voix permettent aujourd'hui d'entendre les textes à lire (Talking Transend, Text-To-Speech System [BUCKLAND, 1986][SPEECH +, 1988]) ou de disposer de calculatrices vocales (The talking Calculator de TSI).

Enfin, les handicapés auditifs et/ou vocaux réussissent à mieux se faire comprendre des entendants, ou peuvent mieux percevoir leur environnement.

Des systèmes de synthèse vocale permettent à des personnes muettes de se faire comprendre au téléphone, elles tapent sur un clavier le texte à prononcer et le système génère les paroles pour leur interlocuteur. Ainsi, le système

portable 'Handivoice' est préprogrammé avec 373 mots, 45 phonèmes (de la langue anglaise), 26 lettres et 16 phrases. Il est conçu en deux versions, l'une pour les personnes avec un handicap intellectuel et l'autre pour les personnes ne possédant qu'une infirmité vocale. Il est accessible via un clavier ou via n'importe quel autre moyen de sélection et permet au handicapé de tenir une conversation complète.

Le problème d'acquisition des sons et du langage pour les déficients vocaux, mal-entendants, atteints de maladie ou victimes d'un accident est grandement résolu par l'intermédiaire des systèmes de rééducation vocale assistée par ordinateur.

Par exemple, le projet Sirène du CRIN permet, d'une part, l'étude en différé des voix pathologiques pour une aide au diagnostic et pour une orientation de la rééducation, et d'autre part, grâce à une contreréaction visuelle, permet la complétion de la boucle audio-phonatoire des jeunes mal-entendants. Ainsi, par l'intermédiaire de systèmes basés sur la théorie du projet et à l'aide de jeux simples, l'enfant prend conscience des types de sons émis vocalement, de la tenue de l'émission d'un son. Il apprend aussi à acquérir des structures rythmiques, à contrôler sa respiration, son souffle et les caractéristiques mélodiques de sa voix.

Un système de ce type est visible en démonstration à Bruxelles, à l'Anspach Center, 4ème étage, IBM-Center.

Scott Instrument a développé un système d'aide à l'apprentissage des langues étrangères dont un module est particulièrement orienté vers l'apprentissage vocal pour des personnes aphasiques ou souffrant de désordre de la communication. [SCOTT, 1988]

La liste des machines mécaniques, électroniques ou informatiques d'aide aux moins valides est longue. A chaque infirmité correspond une interface qui la réduit. Le centre "IBM National Support Center for Persons with Disabilities" a édité, en collaboration avec l'université du Wisconsin, un manuel de référence qui regroupe un très grand nombre d'interfaces pour l'utilisation d'un Personal Computer par une personne infirme. Chacun des produits est décrit brièvement et accompagné de renseignements tels le prix approximatif, l'adresse du constructeur et le type d'interface. [IBM, 1986]

Ce travail est une véritable bible pour tous ceux qui désirent connaître les aides informatiques ou électroniques existantes pour un handicapé.

Désormais, les choses changent, les mentalités évoluent. Un exemple concret donne l'espoir d'une intégration complète des handicapés au sein du circuit "normal" du travail : le 1er septembre 1988, IBM-Belgique ouvre un département destiné à informer tout dirigeant d'entreprise sur des possibilités d'employer des personnes handicapées, grâce aux moyens existants sur le marché. Du point de vue législatif, les gouvernements, eux aussi, s'orientent vers une réelle insertion des personnes handicapées. Par exemple, en Belgique, les administrations sont obligées d'embaucher un certain nombre de handicapés pour des emplois équivalents à ceux des valides, lorsque la fonction le permet.

Après avoir découvert les qualités scientifiques de l'informatique, voilà que nous pouvons lui attribuer une dimension sociale importante. Les utilisateurs "normaux" ne doivent pas être les seuls à réaliser des tâches simplifiées grâce à l'électronique moderne. L'interface homme-machine doit elle aussi s'orienter vers la satisfaction des besoins spécifiques d'individus particuliers dits "anormaux".

4.3 Les handicapés et la reconnaissance vocale

4.3.1 Une mise au point

La technologie vocale peut, elle aussi, apporter un élément de solution aux problèmes d'aide aux personnes handicapées. De nombreux développements et projets dans ce domaine en témoignent.

Le laboratoire d'ingénierie médicale de l'université de Tampere, en Finlande, a développé plusieurs projets ayant pour but la réinsertion du handicapé, avec notamment des systèmes de contrôle d'environnement à commandes vocales [VIT, 1986], et des chaises roulantes à contrôle vocal pour quadriplégiques [RATY, 1988].

Des personnes ayant subi une attaque cérébrale et perdu l'usage d'une parole clairement compréhensible par chacun, peuvent à nouveau avoir des conversations avec des personnes non habituées à leur discours. En effet, les systèmes de reconnaissance vocale sont tous basés sur la reconnaissance de signaux sonores, sans référence à une langue existante particulière, ce qui permet même à des personnes dont la phonation n'est pas compréhensible de pouvoir les utiliser.

En Allemagne, les services de télécommunication pensent à mettre en vente un téléphone totalement "main libre". La composition d'un numéro, la prise d'une communication, l'appel de numéros d'urgence se feraient par commande vocale. [PHILIP-YOUNG, 1986]

Aux Etats-Unis, des constructeurs de systèmes de reconnaissance vocale débloquent des fonds afin de mettre sur pieds des projets d'aide aux handicapés : le système Scott *Voice-Based Learning* (VBLS), des bourses NIH et des contrats avec le département de l'éducation nationale [SCOTT, 1988], des recherches pour l'utilisation d'un PC par un handicapé [IBM, 1987a] [IBM, 1987b], le sponsoring de laboratoires de recherches (*Stanford Research Laboratory*) [KURZWEIL-AI, 1988].

En Angleterre, le docteur Robert I. Damper, du département d'électronique de l'université de Southampton a effectué plusieurs études concernant l'aide aux handicapés par l'intermédiaire de la technologie vocale. [RADAR, 1988], [DHSS, 1988], [DAMPER, 1982]. Dans un article [DAMPER, 1984], il démontre la possibilité de l'exploitation d'un système de reconnaissance vocale automatique bon marché (même peu fiable) pour l'aide aux handicapés physiques. Toujours selon le docteur Damper, l'aide la plus efficace apportée à des tétraplégiques grâce à l'intégration de la voix est celle d'un système de contrôle de l'environnement commandé par la voix.

Efficacité confirmée sur le plan technique. Contrairement à une application informatique générale ou à un logiciel de traitement de textes, un système de contrôle de l'environnement (SCE) possède une applicabilité et un vocabulaire restreints. A l'inverse des télémanipulateurs ou des chaises roulantes, le temps de réponse n'est souvent pas critique. Dans un SCE, le temps de correction d'une erreur

n'est pas susceptible d'engendrer de l'inquiétude. Un taux d'erreur acceptable dans un SCE est souvent trop important dans des applications d'avantage "temps réel" et ne pourrait être toléré.

D'un côté psychologique, le contrôle vocal d'un SCE paraît plus approprié. En effet, donner un ordre à un inférieur ou à un esclave se fait plus naturellement par la voix. Comme un outil de télémanipulation sera perçu plutôt comme une prothèse que comme un esclave, le contrôle vocal ne serait pas idéal. Un contrôle électromyographique serait bien plus approprié. L'utilisateur serait, de ce fait, impliqué dans une boucle de rétroaction pendant le contrôle de "l'engin".

D'un point de vue physiologique, la commande vocale s'adapte facilement à des handicaps très lourds. Elle est peu contraignante et est une interface aseptique idéale. En effet, aucun capteur n'est en contact avec l'utilisateur.

Des SCE à commande vocale existent déjà sur le marché : le *Voice Activated Domestic Appliance System* (VADAS) en étude au département de la santé et de la sécurité sociale anglaise [DHSS, 1988], le Tetravox, un produit français mis au point par le Laboratoire d'Informatique de la Faculté de Médecine de Marseille et le Centre de Réadaptation Fonctionnelle de Valmante : [BENEZET-SAMBUC, 1981], [NEUMEISTER, 1985], [SAINT-SETIERS, 1985], [CREAI, 1987].

4.3.2 Des témoignages

Ces solutions technologiques apportent-elles réellement une aide aux moins valides ? Des handicapés répondent.

Dans de nombreux milieux hospitaliers, paramédicaux et même à domicile, les personnes handicapées utilisent des systèmes informatiques ou informatisés pour l'aide qu'ils procurent. Elles commandent ces systèmes par diverses interfaces (commande à touche unique, bouton-poussoir, micro-rupteur, interrupteur pneumatique répondant au souffle ou à l'aspiration, détecteur de mouvements oculaires, ...) adaptées à leur handicap spécifique.

Selon Christian Meunier, collaborateur bénévole tétraplégique aux projets Aidami, (i.e. projets de SCE associé à un contrôle de logiciel), *"l'intérêt de l'utilisation d'une commande vocale est évident dans le cas d'un handicap très sévère où la possibilité de mouvement est pratiquement nulle, ou bien lorsque les mouvements sont rendus très difficiles en raison des circonstances"* [MEUNIER, 1988]. L'emploi de la reconnaissance vocale permet l'accès à divers systèmes quel que soit le degré d'infirmité ou quelle que soit la possibilité d'accès à ces systèmes. A l'usage des personnes grabataires, pour lesquelles tout déplacement est impossible et dont les mouvements sont limités en fonction de la position choisie (position assise, position couchée sur le dos, sur le ventre ou sur le côté) les mécanismes de commande doivent être réajustés pour chaque nouvelle position adoptée. Or, ces changements de position surviennent le plus fréquemment possible, afin de soulager la personne alitée. Celle-ci risque donc souvent de se retrouver privée de son système parce que son interface de contrôle est hors de portée. La commande vocale apparaît donc comme une solution valable.

Le centre "De Bijtjes - Les Petites Abeilles" de Vlezenbeek accueille des personnes atteintes de myopathie. Suite à des visites de ce centre et à des entretiens avec monsieur Michel Vanliefeland, logopède et responsable du pool informatique de ce centre, nous pouvons faire quelques remarques spécifiques à l'utilisation de la reconnaissance vocale dans le cadre des affections myopathiques.

La myopathie est une affection particulière caractérisée par une diminution progressive de la force musculaire de la personne atteinte. Cette force minimale impose finalement aux handicapés l'usage de manettes de jeux adaptées ou de simples micro-rupteurs.

La rapidité des mouvements est relativement conservée; ainsi, les mouvements ne nécessitant pas une force musculaire importante sont possibles, comme les mouvements des doigts, contrairement aux mouvements de translation des avant-bras. Ceci privilégie des périphériques d'entrée spécifiques. Les myopathes utilisent plus simplement le joystick que la souris, un petit clavier qu'un clavier étendu en surface sur lequel l'accès des touches éloignées est très difficile.

Par contre, l'altération de la parole est beaucoup plus tardive et nettement moins significative sur le plan fonctionnel.

Ces diverses considérations nous amènent à conclure que, pour le type de handicap généré par la myopathie, le pilotage d'un ordinateur par la voix s'avère être l'outil idéal en ce qui concerne l'utilisation de logiciels utilitaires (traitement de texte, gestion de base de données) et de certains logiciels récréatifs.

Les témoignages recueillis soulignent un intérêt important pour ce moyen d'interaction avec un ordinateur.

Ainsi, l'usage d'un éditeur de texte via la commande vocale a soulevé quelques réflexions. La majorité des personnes myopathes ont écrit manuellement. Puis, la maladie évoluant et ne pouvant plus écrire, elles ont eu une capacité d'accès satisfaisante au clavier. Lorsque leur motricité ne leur permet plus de taper sur le clavier, d'autres fonctions, vitales celles-là, sont compromises. Ce contexte fait en sorte que les personnes myopathes ne veulent pas à tout prix réécrire, contrairement à certaines personnes infirmes motrices cérébrales (IMC) ou atteintes de lésions au cordon médullaire ou aux aires corticales. Pour utiliser un éditeur de texte, les systèmes de balayage à touche unique ou à bouton poussoir leur paraît lourd et peu naturel. Il ne leur semble pas opportun de multiplier et de mélanger les systèmes d'accès à l'éditeur. Une solution mixte combinant une introduction vocale (pour l'accès à un ensemble de lettres ou de commandes) et une commande à touche unique (pour le balayage au sein même de l'ensemble) compliquerait un processus déjà par trop difficile.

L'édition de texte via la reconnaissance vocale leur apparaît comme la meilleure solution. La fonction vocale étant une des mieux conservées, ces personnes souhaitent un éditeur de texte vocal complet, c'est-à-dire jusqu'à l'énonciation lettre par lettre des mots à composer. Elles proposent d'associer aux lettres faisant problème un signal vocal permettant leur identification, et ce pour palier à des défauts possibles de reconnaissance des lettres énoncées de façon simple (prononcer "bé", "dé", "pé", "té" les consonnes occlusives b,d,p,t), et pour ne pas utiliser obligatoirement les analogies d'épellation internationale parfois

contraignantes (bravo, delta, papa, tango) et qui posent parfois problème, (mémorisation, prononciation, longueur de certaines analogies, lenteur d'énonciation, ...). Un système ergonomiquement utilisable devrait être mis à l'étude, qui éviterait l'emploi de ces analogies d'épellation internationales (Annexe G).

Les principaux avantages de la reconnaissance vocale relevés auprès des différentes personnes rencontrées concernent tous les handicapés physiques : la puissance, la rapidité d'accès, le confort d'utilisation, un sentiment de sécurité et l'évolutivité de la technique (avantage plus spécifique aux personnes myopathes).

Détaillons-les plus avant :

- La puissance : la reconnaissance vocale et particulièrement les systèmes transparents de contrôle vocal d'un micro-ordinateur peuvent rendre toutes les fonctions des logiciels directement accessibles aux personnes handicapées. Lors de l'utilisation de logiciels du marché, nous sommes très souvent confrontés à la manipulation répétée de combinaisons de touches afin d'accéder à une fonction du logiciel. Les systèmes d'émulation du clavier permettent d'activer en une seule commande vocale un ensemble de touches;
- La rapidité d'accès : contrairement aux systèmes de désignation à balayage (balayage lignes-colonnes ou quadripolaire) associés à une commande unique (touche, contact tout ou rien, ...), la reconnaissance vocale n'entraîne pas de délai d'attente avant que la tâche désirée soit accessible;

- Le confort d'utilisation : il est évidemment plus commode d'allumer une lampe à distance en ne prononçant qu'un mot plutôt que de se déplacer pour actionner l'interface du système de gestion des lumières ou un contact quelconque. De plus, la reconnaissance vocale ne nécessite pas de référent visuel (tableau, écran), surtout si celle-ci est associée à un système de synthèse vocale. Elle n'implique pas la soumission à une vitesse de travail prédéterminée comme sur les systèmes de désignation à balayage. Finalement, elle apparaît comme un système "naturel" aux yeux des utilisateurs handicapés, plus proche d'eux par rapport à leurs possibilités de contrôle;
- Un sentiment de sécurité accru : l'utilisateur peut à tout moment agir sur son système, même dans les situations les plus critiques, quand il se trouve dans l'impossibilité d'atteindre sa commande classique (généralement une commande mécanique);
- L'évolutivité de la technique (pour les personnes myopathes) : l'accès direct au clavier diminuant avec l'évolution de la maladie, la reconnaissance vocale pourra prendre progressivement en compte la gestion des touches ou des séquences de touches qui commandent les fonctions des logiciels. Demeurant pour un temps la voie privilégiée de l'édition de texte, le clavier ne sera remplacé par la reconnaissance vocale que lorsque celui-ci ne sera plus utilisable.

Notons, que de ces cinq avantages de la reconnaissance vocale, la rapidité d'accès et le confort d'utilisation sont les plus mentionnés par les personnes rencontrées.

Une diversification des types d'interface de contrôle d'un ordinateur signifie, comme dans le cas du système Aidami, l'ouverture d'un SCE intégrant l'utilisation de logiciels, à d'autres utilisations, professionnelles, commerciales. L'accès à des logiciels plus sophistiqués est possible précisément grâce à la combinaison de différents types de commande dont la reconnaissance vocale (CAO, comptabilité).

Après avoir décrit la situation des handicapés par rapport à l'informatique, nous allons nous arrêter à un système d'aide à ces personnes : le projet Aidami.

4.4 Le projet Aidami

Pouvoir téléphoner, recevoir des appels téléphoniques, allumer la télévision, dactylographier un texte pour un ami et qui plus est, utiliser un programme comme D-base pour entretenir ses fichiers, voilà ce que le projet Aidami offre à la personne handicapée : une autonomie retrouvée.

Le projet Aidami (Aide à partir de Microprocesseurs) est un projet qui, initialement, a été développé par la faculté Polytechnique de Mons, afin d'aider une personne atteinte de sclérose en plaques à conserver un moyen de communication avec son entourage par l'intermédiaire d'un éditeur de texte fonctionnant uniquement à l'aide d'un bouton poussoir.

Il a rapidement évolué vers un système d'aide et de contrôle de l'environnement pour personne dont la mobilité est faible, plus particulièrement pour des personnes atteintes de tétraplégie. Mais, basé sur une technologie fermée, le projet se devait de s'orienter vers une utilisation de matériels conventionnels et une architecture plus ouverte. La partie logicielle de la nouvelle génération du projet a été développée par une équipe dirigée par le Professeur J. Ramaekers des Facultés Notre Dame de la Paix à Namur.

Ce nouveau projet Aidami (dont le sigle pourrait désormais signifier Aide à partir d'un Micro-ordinateur) caractérisé par l'emploi de matériel standard, en l'occurrence d'un micro ordinateur du type Personal Computer d'IBM, réunit les mêmes qualités que le projet précédent mais permet également à la personne handicapée qui l'utilise d'employer tous les logiciels existant sur le marché.

Aujourd'hui, suite aux travaux de cette équipe [ANDRE et al., 1987], travaux complétés par le mémoire de Philippe Thirionet [THIRIONET, 1987], une personne tétraplégique a vu ses besoins en "autonomie informatisée" résolus.

La dernière version d'Aidami s'oriente vers l'utilisation de boutons-poussoirs, de *joysticks*, de *trackballs* et de souris, interfaces simples de saisie de données. La reconnaissance vocale ne pourrait-elle pas, elle aussi, jouer un rôle dans la manipulation d'un tel logiciel ?

4.4.1 Les fonctionnalités du système

Le système Aidami a pour but de remplir différentes fonctionnalités. Ces fonctionnalités sont regroupées en "tâches". Ces tâches sont dénommées "COURRIER", "TELEPHONE", "RELAIS", "TELECOMMANDE", "OUTILS" et "LOGICIEL". Une "SORTIE" du système a aussi été prévue, bien que sans intérêt pour la personne handicapée. Chacune des tâches possède des fonctions précises.

- La tâche "COURRIER" regroupe essentiellement les fonctions d'édition et d'impression de textes, de gestion de fichiers (fichiers de travail, fichier-agenda, fichier-annuaire, notes, ...).

- La tâche "TELEPHONE" donne la possibilité à l'utilisateur de donner ou de recevoir un coup de téléphone, ainsi que de prendre des notes. Cette tâche présente trois possibilités pour effectuer un appel téléphonique. L'utilisateur peut composer un nouveau numéro à l'aide d'un cadran téléphonique, sélectionner un numéro dans l'annuaire ou refaire le dernier numéro composé.

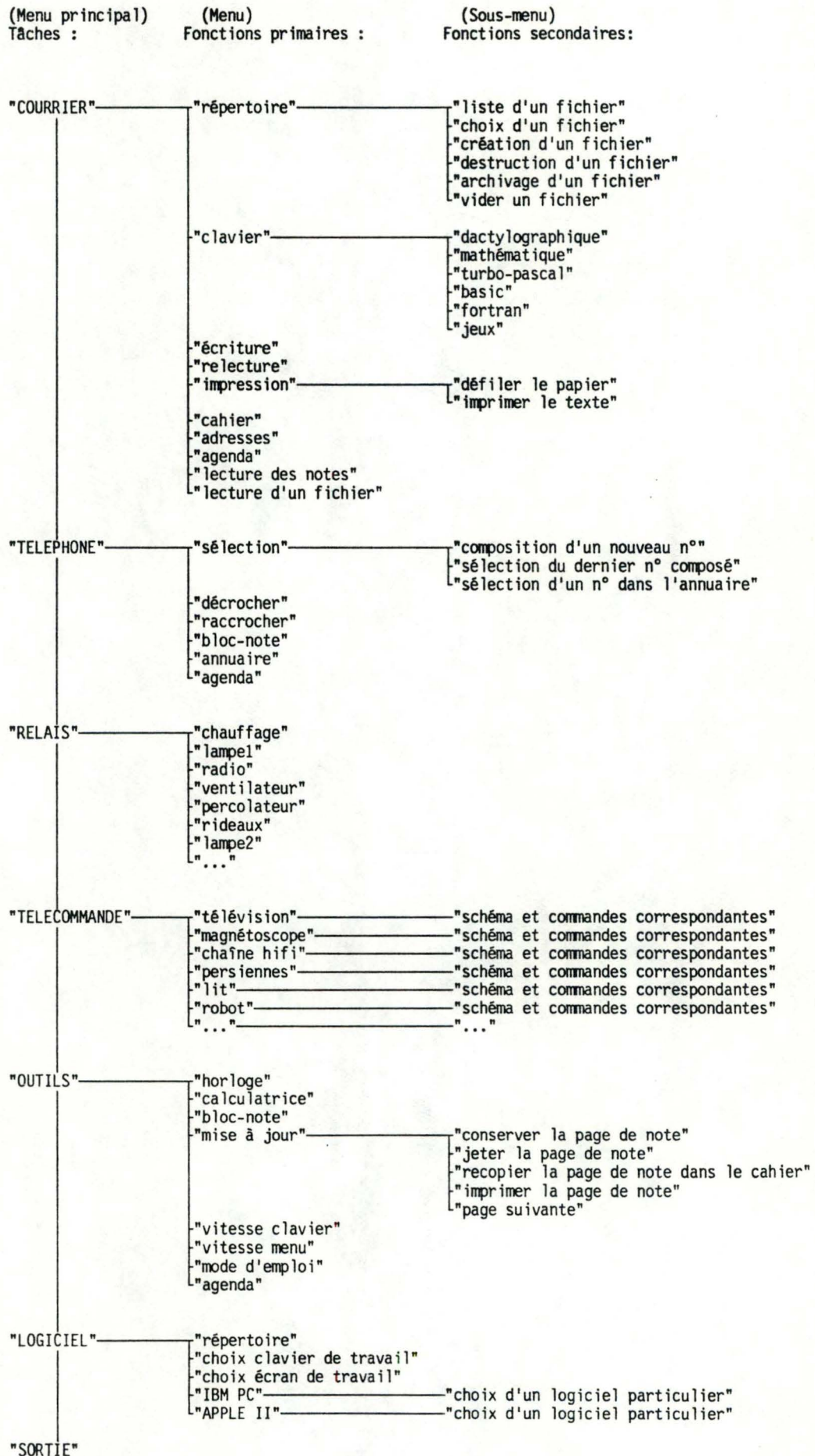


Figure 4.1 : Les fonctionnalités du système Aidami

- La tâche "RELAIS" permet d'agir sur l'un des appareils électriques connectés au bornier. Il suffit de sélectionner l'appareil pour allumer ou éteindre celui-ci.

- La tâche "TELECOMMANDE" regroupe les différents appareils commandés par télécommande. Le dessin de la télécommande associée à l'objet est présenté à l'écran et l'utilisateur sélectionne sur ce schéma la fonction désirée.

- La tâche "OUTILS" groupe des fonctions utilitaires comme l'affichage de l'heure, l'affichage d'un mode d'emploi, la mise à jour des notes prises, le choix de vitesse de balayage des menus.

- La tâche "LOGICIELS" permet à l'utilisateur de "quitter" l'environnement Aidami afin d'utiliser n'importe quel logiciel compatible avec l'ordinateur.

Les différentes fonctionnalités du système Aidami s'enchaînent selon la figure 4.1. Chacune des tâches correspond à un item du menu déroulant principal. Chacune des fonctions primaires est représentée dans un menu et chacune des fonctions secondaires dans un sous-menu du menu précédent.

Les fonctionnalités du système Aidami décrites, nous allons maintenant aborder l'architecture physique et logicielle du système.

4.4.2 L'architecture physique et logicielle du système

Le système Aidami de base (version de [ANDRE et al., 1987]) est construit autour d'un Personal Computer. Le système essentiel d'entrée de données combine la présentation d'une suite de menus de désignation par balayage (ou menu déroulants) et l'action d'un bouton-poussoir. L'interface principale de sortie est constituée d'une carte d'interface parallèle insérée dans le micro-ordinateur qui contrôle un boîtier gérant tous les accès aux périphériques externes.

Les menus déroulants sont composés par un certain nombre d'items qui correspondent chacun soit à une action à effectuer sur le système (sauver le fichier, allumer une lampe, ...), soit à un sous-menu à afficher (l'item "répertoire" du menu "COURRIER" permet d'afficher le sous-menu "REPERTOIRE"). Chaque sous-menu est lui-même un menu déroulant.

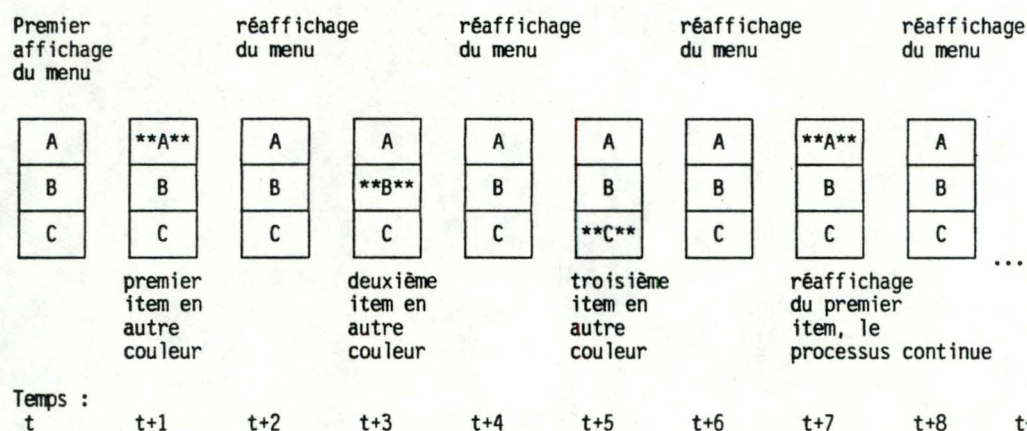


Figure 4.2 : Principe du menu déroulant comme implémenté dans la version de base de Aidami

Pour réaliser l'effet de balayage, les items d'un menu sont affichés chacun l'un à la suite de l'autre en une "autre" couleur (par exemple : reverse video). C'est la rapidité d'affichage qui donne l'illusion du balayage. La figure 4.2 présente le principe sous-jacent à ce système de menu. Notons que ce principe de balayage fonctionne pour des menus dont les items sont représentés aussi bien verticalement qu'horizontalement.

La sélection réelle est réalisée via un bouton-poussoir qui, une fois enfoncé, génère une interruption logicielle. Cette interruption logicielle est capturée par le programme Aidami qui, selon l'item "actif" à ce moment, va déterminer la tâche à effectuer, l'impulsion à donner au boîtier central ou l'affichage du sous-menu correspondant (se reporter à la section décrivant les fonctionnalités du système Aidami pour plus de détails au sujet de l'agencement des menu en sous-menus).

Pour certaines tâches, des interfaces de sortie particulières sont indispensables (télécommandes, téléphone, bornier). Ces interfaces sont connectées au boîtier central Aidami qui est lui-même contrôlé par une carte insérée dans le micro-ordinateur et commandée par le programme Aidami.

Pour la tâche "TELECOMMANDE", fonctions "télévision" et "magnétoscope", une télécommande à diode infrarouge envoie les codes nécessaires à la conduite du téléviseur et de l'enregistreur video. Ceci implique bien entendu que le poste de télévision et le magnétoscope soient tous deux commandables à distance et que l'on connaisse les codes infrarouges auxquels ils réagissent.

Pour la tâche "TELEPHONE", un appareil téléphonique "main libre" est connecté au boîtier central. Sur les instructions de la carte, le boîtier décroche le téléphone, compose un numéro, raccroche l'appareil.

Pour la tâche "RELAIS", le boîtier central peut commander l'ouverture ou la fermeture de relais statiques de puissance contenus dans un bornier. Ces relais de puissance permettent de contrôler des appareillages 220 volts (mise en route et arrêt).

L'architecture du système Aidami décrite, nous allons souligner quelques points perfectibles. Le projet de base évoluée, une nouvelle équipe s'est mise au travail.

4.4.3 Remarques et orientations futures

Afin de permettre un contrôle plus aisé de la carte d'interface parallèle et une manipulation plus simple des interruptions générées par le bouton-poussoir, les interruptions standards du Bios du Personal Computer ont été modifiées.

Cette modification a fait en sorte que le système Aidami ne soit pas totalement compatible avec l'architecture des IBM-PC, ce qui implique que le système n'est pas réellement facilement utilisable par n'importe qui. Citons par exemple le fait que le clavier n'est pas accessible durant l'étape de contrôle de l'environnement.

Pour Christian Meunier, collaborateur tétraplégique connaissant tous les pièges insidieux du programme, la version du système Aidami qui tourne chez lui est parfaite. Ses besoins en contrôle de l'environnement sont réalisés. Il peut utiliser un grand nombre de logiciels standards et ainsi répondre à ses besoins informatiques. Le téléphone "main-libre" commandé par le système Aidami lui permet de contacter ses amis et de recevoir des appels téléphoniques. Bref, il a retrouvé une certaine autonomie. Mais le système reste un remarquable prototype, avec de grandes qualités et quelques défauts.

Aujourd'hui, une équipe s'est remise au travail. Son objectif : *"faire du projet Aidami, un produit Aidami, standard qui peut être acquis par n'importe qui"*. Une nouvelle politique d'implémentation a été décidée : *"utiliser au maximum des éléments existant sur le marché et totalement compatibles ou rendus compatibles avec l'architecture d'un Personal Computer"*. La nouvelle architecture du système va déboucher sur un ensemble de "modules" indépendants entre eux et qui réaliseront chacun différentes tâches. Par exemple, un des modules centraux, celui d'acquisition des données, pourra recueillir des instructions via n'importe quelle interface d'entrée, aussi bien un clavier géant qu'un interrupteur au souffle, ou ... un système de commande vocale. L'intégration de n'importe quelle interface de contrôle sera alors réalisée.

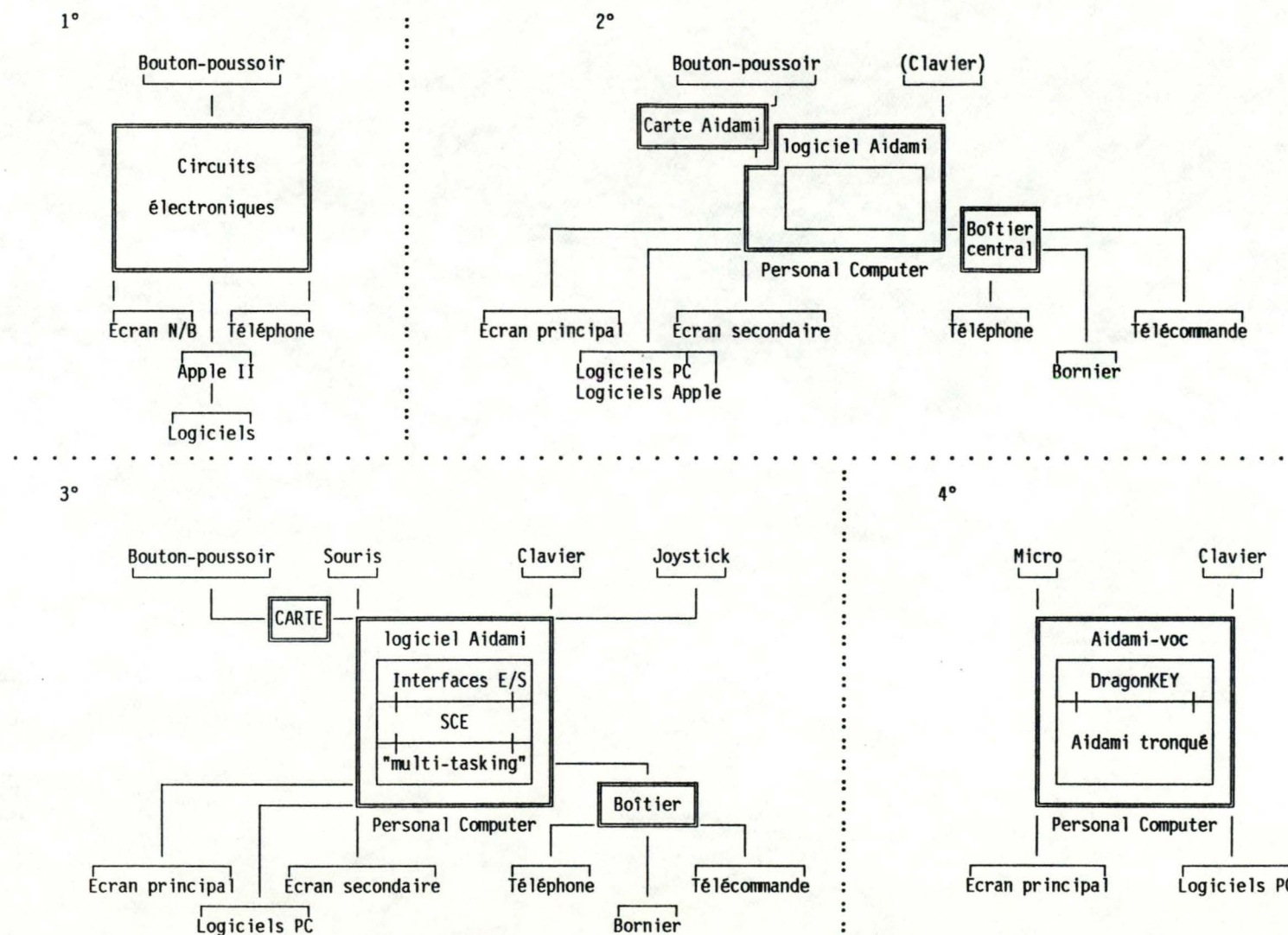
Une fois tous les modules mis au point, le système pourra être acquis "morceau par morceau", selon les besoins de l'utilisateur.

Les différentes versions du projet Aidami sont visibles sur la figure 4.3.

- 1° correspond à la version réalisée exclusivement sur base de circuits électroniques par la faculté polytechnique de Mons.
- 2° représente la version développée par [ANDRE et al., 1987], réalisée sur un IBM Personal Computer.
- 3° désigne la version de P. Thirionnet [THIRIONET, 1987] (amélioration de 2°, extension des contrôles à d'autres interfaces E/S, modularisation et implémentation basée sur une certaine multiprogrammation).
- 4° correspond à la représentation du système Aidami-voc, prototype de l'application vocale développé dans le cadre de ce mémoire.

Mais, la voix peut-elle être intégrée à un SCE du type Aidami ? Nous allons montrer dans la section suivante que la voix se prête particulièrement bien à la conduite de la version de base de Aidami, mais pour cela, il est nécessaire de transformer quelque peu le SCE de base pour que la commande vocale soit totalement performante.

Figure 4.3 : Récapitulatif des versions du projet Aidami



4.5 L'intégration de la voix au projet Aidami

"La reconnaissance vocale commence à être appliquée pour aider ceux qui sont sévèrement handicapés physiquement comme les quadriplégiques ou ceux avec des blessures au cordon médullaire. Pour certaines de ces personnes, la parole peut être leur seul moyen de communication, ainsi les systèmes vocaux peuvent être d'une importance vitale pour eux. Cela peut prendre du temps d'apprendre à parler correctement à un ordinateur; une forte motivation est essentielle pour l'utilisateur. Celle-ci manque parfois aux autres utilisateurs qui peuvent choisir un autre moyen de dialogue avec la machine."

[PHILIP-YOUNG, 1987] p 19

Le système Aidami de base, conçu pour fonctionner sur un IBM Personal Computer, permet grâce à un bouton-poussoir et par l'intermédiaire de menus déroulants de choisir une tâche à réaliser. C'est ce principe de sélection par balayage qui a retenu notre attention. Si l'on considère ces menus comme des entités propres, chacun d'eux correspond à une action précise ou à un choix précis à effectuer. Cette architecture convient parfaitement à un système de commande unique associée à chacune des actions à effectuer et ainsi à la reconnaissance vocale de mots-clés isolés.

Afin d'y intégrer une "commande vocale" et d'en faire une "application vocale" dans les termes définis au chapitre III, nous avons choisi la version du programme Aidami développée par P. André et par B. Bellofatto [ANDRE et al., 1987], principalement pour deux raisons. La première est que cette version du programme Aidami fonctionne dans un environnement réel, ce qui pourra nous aider si des choix d'implémentation

sont à faire (fonctionnalités à ajouter ou à supprimer, choix des mots du vocabulaire, ...). La seconde est qu'en choisissant cette version, nous ne serons pas confrontés à des tâches incomplètes ou à des routines de manipulation à mettre au point.

L'intégration de la voix dans le système Aidami s'est effectuée en plusieurs étapes dont voici une description succincte :

- La première étape de conception du système a été une étape de *destruction*. Nous avons éliminé les conséquences réelles des fonctionnalités du programme, pour ainsi démontrer plus facilement, sur un système non équipé de relais statiques de puissance (pour la commande d'appareillage électroménager) et non équipé de télécommande (pour un téléviseur ou une chaîne Hi-Fi) ou de téléphone "main-libre", l'utilité de l'usage de la voix pour commander différentes "pseudo-tâches" (tâches sans réalisation concrète).
- La seconde étape a été une étape d'*inventaire*. Nous avons, en effet, recensé toutes les tâches du système ainsi que leur déroulement afin de pouvoir construire un vocabulaire de base.
- La troisième étape a été caractérisée par une *construction*. Nous avons construit un système de menus semblable à celui déjà développé mais qui laissait le clavier libre de travailler. Ainsi, il nous est possible de généraliser l'utilisation de la reconnaissance vocale en combinaison d'un autre type d'interface d'entrée/sortie.
- La quatrième étape est une étape de *synthèse*. Nous avons intégré le principe de contrôle de la carte de reconnaissance dans toutes les tâches du programme même afin d'aboutir à un "logiciel vocal".

Le but poursuivi est de construire une application informatique correspondant au système Aidami, mais répondant aussi bien au clavier qu'à la voix, afin de démontrer la facilité d'emploi d'une véritable "mise en pratique" vocale en opposition avec une "application" de la technologie vocale à un programme déjà établi. Cette "mise en pratique" vocale serait l'étape préliminaire à une extension possible du projet Aidami à d'autres interfaces d'entrée/sortie.

Ce logiciel-prototype, nous l'appellerons Aidami-voc.

4.5.1 Première étape : destruction

La première étape de développement de l'application Aidami-voc a été de la rendre compatible avec un micro-ordinateur Personal Computer non équipé des différents périphériques de contrôle de l'environnement. Nous avons donc dû modifier les sous-routines d'initialisation et supprimer ou détourner tous les appels aux périphériques.

Grâce à ces modifications, le clavier est redevenu utilisable, bien qu'aucune instruction d'entrée au clavier n'apparaît dans les sources. Le programme pouvait, enfin, s'exécuter sans conduire à un crash de la machine (situation courante si la carte d'interface parallèle Aidami n'est pas installée dans le micro-ordinateur).

4.5.2 Deuxième étape : inventaire

Comme nous l'avons vu dans le chapitre III, le système de reconnaissance vocale VoiceScribe 1000 nécessite une grammaire décrivant exactement les termes à employer et leur enchaînement les uns par rapport aux autres. La deuxième étape a été de recenser les intitulés des différentes tâches présentes dans le système Aidami.

C'est ainsi que nous avons construit l'arbre d'enchaînement des menus et sous-menus, représenté par la figure 4.1 de la section précédente.

Cet "arbre de déduction logique" nous aidera dans la réalisation d'une grammaire correcte pour le logiciel vocal désiré.

4.5.3 Troisième étape : construction

Dans la version du programme Aidami sur laquelle nous basons pour développer l'application vocale Aidami-voc, chaque menu est du type "à désignation par balayage", c'est-à-dire que pour sélectionner la tâche ou l'action à effectuer, il suffit d'agir adéquatement sur l'interface mise à la disposition de l'utilisateur handicapé (dans le cas qui nous préoccupe, appuyer sur un bouton-poussoir du type micro-contact) lorsque le balayage passe sur le nom de l'action à effectuer.

Dans la mesure où nous développons un prototype vocal, ce système de balayage est inutile. Il suffit de donner vocalement le nom de la tâche à effectuer pour que celle-ci se réalise.

Un premier pas dans la troisième étape consiste donc à supprimer ce balayage, devenu intempestif. Pour ce faire, comme le balayage n'était pas structuré dans une routine particulière, nous avons dû parcourir ligne après ligne, le code source pour supprimer cet effet de balayage.

Permettre une simple commande vocale des tâches d'Aidami n'est pas le seul but poursuivi. Nous voudrions construire une véritable application vocale mais dont le clavier est toujours en fonction. Un second pas consiste alors à trouver un artifice le permettant.

Afin de commander via le clavier les différentes tâches du système Aidami, nous avons donc associé à chacun des items de menu une lettre particulière désignant la touche déclenchant l'action correspondant à cet item. Il suffit donc de pousser cette touche pour que la tâche s'effectue ou que le sous-menu s'affiche.

Par exemple, la tâche "SORTIE" est associée à la touche 'S'. Il suffit donc d'appuyer sur la touche 'S' pour sortir de l'application. Les associations des touches et des items de menu sont données in extenso en annexe.

4.5.4 Quatrième étape : synthèse

Lors de la conduite de logiciels du marché (Lotus, D-base, ...), le système de reconnaissance vocal peut entrer en désynchronisation vis-à-vis des commandes du logiciel.

Afin d'effectuer une tâche quelconque, si l'utilisateur commence une séquence d'action en utilisant la commande vocale puis continue cette séquence en utilisant le clavier, le système de reconnaissance ne sera pas averti de l'évolution de la tâche. Dans ce cas précis, le vocabulaire "actif" correspondra toujours à un ensemble de commande "actives" précédentes et non pas à la nouvelle situation.

La quatrième étape consiste donc en la construction (dans le programme Aidami-voc) de la "synchronisation" entre la conduite de l'application par le clavier et par la commande vocale.

Cette synchronisation se fera simplement. Chaque menu affiché enverra un code de contrôle à l'interface vocale pour lui signaler à quel menu l'application se situe. Quoiqu'il arrive lors de l'accès à un nouveau menu, le système de reconnaissance vocale sera averti du déroulement des opérations ainsi aucune désynchronisation entre le vocabulaire actif (i.e. les mots susceptibles de reconnaissance) et les noms des items du menu actif ne pourra être détectée.

Finalement, la grammaire utilisée est une grammaire simplifiée dont le texte est en annexe.

4.6 Perspectives d'avenir

Au terme de ce mémoire, mentionnons encore les différentes orientations selon lesquelles les recherches pourront être poursuivies sur base du travail déjà réalisé.

La première orientation est de redessiner un système Aidami plus standard, plus modulaire, sur lequel pourrait se greffer n'importe quelle interface d'entrée/sortie. Il est à noter que ce nouveau développement du projet est déjà à l'étude sur base du mémoire de P. Thirionet [THIRIONET, 1987].

Afin d'intégrer parfaitement la voix au projet Aidami et ainsi développer une application complète (non tronquée) d'aide aux handicapés, nous aurions dû réaliser une nouvelle analyse des fonctions proposées par Aidami et ainsi obtenir un système plus performant et mieux conçu. Mais, l'objet du mémoire étant différent, il nous a paru plus important de démontrer les qualités de la reconnaissance vocale dans une optique d'aide informatique aux handicapés.

Dans cette même optique de développement informatique et suite aux réflexions sur le système VoiceScribe 1000, il serait intéressant de réaliser un système Aidami X qui, résident dans la mémoire de l'ordinateur, agirait avec le même type de fonctionnalités que *DragonKEY*. Présence invisible, en retrait de chacun des logiciels utilisés, cet Aidami X n'apparaîtrait réellement que lorsque l'utilisateur lui ferait explicitement appel. Comme le représente la figure 4.4, l'optique sous-jacente à cet Aidami X serait différente : ne pas voir en Aidami X un programme qui, toujours actif, pilote d'autres logiciels et gère le contrôle de l'environnement (4.4a) mais plutôt comme un programme caché qui aide au pilotage d'autres programmes dont un logiciel particulier de contrôle de l'environnement (4.4b).

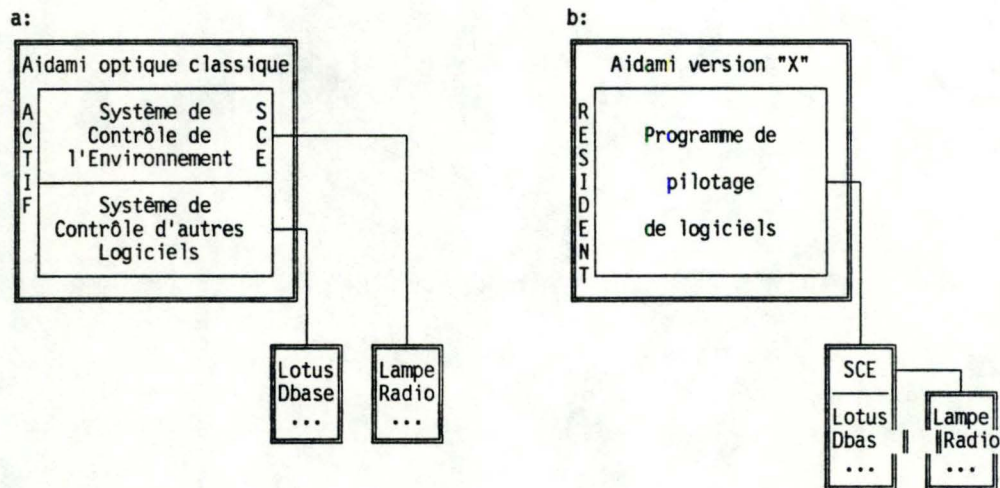


Figure 4.4 : Optique de construction de Aidami (a) et de Aidami X (b)

La seconde orientation est de réaliser une étude plus complète de l'aide informatique aux handicapés. Cette étude pourrait s'effectuer en collaboration avec des organismes comme le Psynha du département de psychologie de la faculté de médecine des FUNDP à Namur, en organisant des rencontres avec des personnes atteintes de divers handicaps, en évaluant leurs besoins réels, en proposant des tests "sur le terrain" de prototypes réalisés, ... Cette étude aiderait à démontrer et à cerner plus systématiquement l'aide que peuvent apporter l'outil informatique et les systèmes comme ceux développés dans le projet Aidami.

Dans cette double perspective, nous soulignons ici les principales difficultés rencontrées durant la réalisation de ce mémoire dans l'espoir de renseigner clairement les personnes intéressées par l'approfondissement du projet.

La phase de documentation et de recherche bibliographique a nécessité beaucoup de temps. Les bases de données de références littéraires ne possèdent pas de mot-clé pour "aide informatique aux handicapés". La reconnaissance vocale elle-même ne fait plus l'objet de nombreuses parutions comme c'était le cas dans les années 70. Seuls des articles scientifiques de pointe peuvent être trouvés relativement facilement, mais pour une utilisation pratique de la reconnaissance vocale, point n'est nécessaire de connaître les détails de mise au point du dernier projet Z du professeur X.

Nous avons aussi à déplorer l'inertie des pouvoirs publics et privés devant *le problème "handicapé"*.

Dans une optique sociale, le continent devrait peut-être prendre exemple sur les îles britanniques. En effet, afin de recueillir des renseignements sur l'utilisation possible de la reconnaissance vocale pour l'aide aux handicapés dans différents milieux, nous avons écrit à de nombreuses institutions, sociétés, centres et associations belges, françaises et anglo-saxonnes. Peu de réponses réellement intéressantes nous sont parvenues des pays continentaux. Par contre, nos contacts anglais n'ont pas hésité à nous informer sur leurs activités et ont même signalé l'action d'Aidami dans le domaine de la parole à des chercheurs finlandais travaillant sur un sujet semblable, chercheurs avec qui nous avons établi un échange de références bibliographiques et d'expériences pratiques de la reconnaissance vocale.

D'un point de vue plus industriel, les constructeurs contactés ont en général répondu favorablement à l'appel d'offre qui leur était lancé. La documentation nous a permis d'effectuer un choix judicieux parmi les différents systèmes de reconnaissance vocale sur le marché.

Pour devenir acquéreur d'un appareil de reconnaissance vocale et ainsi l'étudier "en pratique", il est néanmoins obligatoire de posséder les fonds nécessaires à son achat. Après de vaines démarches administratives, nous nous sommes finalement tournés vers le secteur privé pour découvrir un mécène qui nous a permis de continuer le développement du mémoire dans l'axe désiré.

4.7 En résumé

Dans ce chapitre, nous nous sommes attardés sur les possibilités pratiques d'aide aux handicapés offertes par la technologie informatique en général et par la technique de la reconnaissance vocale en particulier.

Pour cela, nous avons tout d'abord souligné les problèmes spécifiques rencontrés par des les handicapés dans l'utilisation de l'informatique (problèmes de manipulation de clavier, ...)

Ensuite, nous avons montré que la reconnaissance vocale pouvait apporter une solution à ce type de problème.

Enfin, nous avons quelque peu décrit le projet Aidami, projet d'aide aux handicapés sensori-moteurs, à partir duquel nous avons développé le système Aidami-voc, prototype d'une application vocale d'aide aux personnes handicapées.

Conclusions

Conclusions

"Si quelqu'un s'évanouit, on réclame de l'eau, les gouttes de Hoffman, l'eau de Cologne; mais quand un homme s'abandonne au désespoir, il convient de dire : trouvez-lui une possibilité; trouvez-en une; c'est la seule voie de salut. Est-elle donnée : le désespéré retrouve la respiration, reprend vie; car sans possibilité, l'homme ne peut pas, pour ainsi dire respirer..."

[KIERKEGAARD, 1850]

Dans ce mémoire, nous avons effectué une synthèse globale du problème de la reconnaissance vocale. Ensuite, en étudiant son intégration dans un ensemble micro-informatique, nous avons cherché à savoir si la reconnaissance vocale n'était pas un gadget supplémentaire et superflu dans l'aide aux handicapés.

La réponse est négative. La reconnaissance vocale n'est pas un simple gadget. D'après les témoignages recueillis et les conclusions avancées, les personnes handicapées souhaitent un système d'aide très fonctionnel et qui leur offre l'ensemble d'interfaces d'entrée/sortie le plus général possible. Vis-à-vis d'une intégration dans un système d'aide aux non-valides, la reconnaissance vocale et plus particulièrement la commande vocale apparaît comme une interface valable et non pas comme un luxe ludique. La reconnaissance vocale se pose donc comme une interface au même titre que le bouton-poussoir ou l'interrupteur pneumatique.

Le travail développé dans ce mémoire n'est pas une fin en soi et il doit pouvoir aider à l'élaboration d'un ensemble plus fonctionnel, plus modulaire, plus ergonomique de logiciels d'aide aux handicapés dans lesquels pourrait s'intégrer un module de dialogue dont une des fonctionnalités serait spécifiquement conçue pour la reconnaissance vocale. En outre, sur base de ce mémoire, une étude pourrait se développer afin d'évaluer l'aide qu'apporterait de manière générale l'informatique aux handicapés.

Pour conclure, après toutes ces considérations techniques, qu'il nous soit permis enfin d'insister sur l'apport humain et social à la formation d'un étudiant en informatique, qu'a constitué la découverte du monde des personnes handicapées.

Bibliographie

[ANDRE et al., 1987]

P. ANDRE, J. DUTERNE, B. BELLOFATTO, J.M. LASSOIE : *Le Système Aidami Mons-Namur; Contrat cadre spécial temporaire, CST n°30319, FPM & FNDP, (1986-1987).*

[Asimov, 1950, "Robbie"]

I. ASIMOV : "Robbie", in *I, ROBOT*; Doubleday & Co (1950).

[BARD, 1988]

Extraits de British Database on Research into Aids for the Disabled, (Bard) Enregistrements n° 77, 353, 583, 623, 624, 993.

[BECKMAN, 1982]

G. BECKMAN : *As speech recognition products improve , office application appear*; Electronic Design, SS5-SS6, (June 1982)

[BENEZET-SAMBUC, 1981]

P. BENEZET & R. SAMBUC : *Réalisation d'un système de contrôle de l'environnement du tétraplégique utilisant un micro-ordinateur*; Annales de médecine physique XXIV, n°2, pp 167-168 (1981).

[BRUCE-GIBSON, 1984]

I.P.C. BRUCE & D.L. GIBSON : *Using automatic speech recognition to control new telephone services; Proceedings American Voice I/O System Application (AVIOS) Conference, Arlington, Virginie, (11-13 sept. 1984).*

[BUCKLAND, 1986]

B. BUCKLAND : *Text-to-Speech System Offers Unlimited Potential*; Santa Clara County Business, p 9, (July 1986).

[C.H.I., 1988]

advertisement of proceedings C.H.I. (1988).

[CHANDELON-WARNANT, 1987]

M. CHANDELON & G. WARNANT : *Modélisation d'une application interactive*; mémoire présenté en vue de l'obtention du titre de licencié et maître en Informatique, FUNDP, Namur, (1982).

[CLARK-ROEMER, 1977]

J.A. CLARK & R.B. ROEMER : *Voice controlled Wheelchair*; Archives of Physical Medicine and Rehabilitation, 58, pp 169-175, (1977).

[COHEN-GRAUPE, 1980]

A. COHEN & D. GRAUPE : *Speech recognition and control system for the severely disabled*; Journal of Biomedical Engineering 2, pp 99-107, (1980).

[COMPAID, 1988]

Correspondance avec Computer Aid for Speech Impaired and Disabled People

[Coler, 1977]

C.R. COLER, R. PLUMMER, E.HUFF, M. HITCHCOCK : "Automatic speech recognition research at NASA Ames Research Center" in R. BREAUX, M.CURRAN & E.HUFF (Editors), *Voice Technology for interactive realtime command/control systems applications*, Moffet Field, Californie, (Décembre 1977).

[Cotton, 1983]

J.C. COTTON, M. McCAULEY, R. NORTH & M. STRIEB : *Development of speech input/output for tactical aircraft*, Wright-Patterson AFB, Ohio, (1983).

[CREAI, 1987]

R. SAMBUC : *TETRAVOX, la reconnaissance automatique au service des handicapés*; Actes du colloques Informatique et Handicap en institutions et services spécialisés, C.R.E.A.I., (26-27 mai 1987).

[CREASEY, 1982]

G.H. CREASEY : *Voice controlled aids for the handicapped*; Proceedings, World Congress on medical physics and biomedical engineering, Hamburg, (septembre 1982).

[DAMPER, 1182]

R.I. DAMPER : *Speech technology - implications for biomedical engineering*; Journal of Medical engineering & Technology, Vol. 6, n°4, pp 135-149, (1982).

[DAMPER, 1984]

R.I. DAMPER : *Voice-Input Aids for the Physically Disabled*; International Journal of Man-Machine Studies, Vol. 21, n°6, pp 541-553, (1984).

[DAMPER-DABBAGH, 1983]

R.I. DAMPER & H.H. DABBAGH : *Voice-input word-processor aid for the physically disabled*, Conference on Microelectronics in Health Care, Leeds, (June 1983).

[DAMPER-PURSWANI, 1982]

R.I. DAMPER & V.N. PURSWANI : *Voice-input environmental control for the physically disabled*; Proceedings, World Congress on medical physics and biomedical engineering, Hamburg, (septembre 1982).

[DAVIS K.H., R. BIDDULPH and S. BALESHEK, 1952]

K.H. DAVIS, R. BIDDULPH & S. BALESHEK : *Automatic recognition of spoken digits*; J. Acoust. Soc. Amer.; Volume 24, p 637-642, (1952).

[DEMORTIER-LUCAS, 1984]

G. DEMORTIER & A. LUCAS : *Physique générale*; Cours de première candidature en Sciences Mathématiques, FUNDP, Namur, (1984).

[DHSS, 1988]

correspondance avec le Department of Health and Social Security

[DODDINGTON-SCHALK, 1981]

G.R. DODDINGTON & Th.B. SCHALK : *Speech recognition : turning theory to practice*; IEEE Spectrum, pp 26-32, (Sept. 1981)

[DUCHREN, 1985]

D. DUCHREN : *Review of user applications for a small user voice store and forward system*; Proceedings American Voice I/O System Application Conference (AVIOS), Arlington, Virginie, (11-12 sept. 1985).

[DUDLEY-BALASHEK, 1958]

H. DUDLEY & S. BALASHEK : *Automatic Recognition of phonetic patterns in speech*; Journal of the Acoustical Society of America, 30; pp 721-739, (1958).

[DUSEK-SCHALK-MCMAHAN, 1983]

L. DUSEK, TCB. SCHALK & M. MCMAHAN : *Voice recognition joins speech on programmable board*; Electronics Week 56 (8), (1983).

[ERMAN L.D., 1980]

L.D. ERMAN, F. HAYES-ROTH, V.R. LESSER & D.R. REDDY : *HEARSAY-II Speech-Understanding System : Integrating Knowledge to Resolve Uncertainty*; Computing Surveys, Volume 12, n°2, p 213-253, (1980).

[Ferretti-Cinare, 1983, pp 15-24]

M. FERRETTI & F. CINARE : *Synthèse, reconnaissance de la parole*; EdiTests, Paris , (1983).

[Forgie J.W. & Forgie C.D., 1959]

J.W. FORGIE & C.D. FORGIE : *Results obtained from a vowel recognition computer program*; J. Acoust. Soc. Amer.; Volume 31, p 1480-1489, (1959).

[FUEGE, 1978]

R.L. FUEGE & C.W. GEER : *Integrated applications of automated speech technology*; (Rapport ONR-CR213-58-IAF), Seattle, Washington, (1978).

[GLENN et al., 1976]

J.W. GLENN, K.H. HILLER & M.T. BROMAN : *Voice terminal may offer opportunities for employment to the disabled*; American journal of occupational therapy, 30, pp 309-313, (1976).

[GODWIN, 1982]

A.M. GODWIN & B.B. WYATT : "Voice interactive system development program" in *NAECON Proceedings*, IEEE press, New York, (1982).

[HAGEN, 1985]

B. HAGEN (editor) : *Closing the Gap*; revue bi-mensuelle, Henderson, Minnessota, (Avril/mai 1985).

[HANSON-WIATROWSKI, 1985]

G. HANSON & C.A. WIATROWSKI, *Electronics WEEK* 58 (3), pp 45-47, (1985).

[HARRIS, 1985]

S.D. HARRIS (editor) : *Proceedings of the Voice Avionics Conference*; Warminster, (1985).

[HATON J.P. & PIERREL J.M., 1976]

J.P. HATON & J.M. PIERREL : *"Organization and Operation of a Connected Speech Understanding System at Lexical Syntactic and Sementic Level"*, Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing 76; Philadelphia, (1976).

[Harris, 1980]

S. HARRIS (editor) : *Voice Interactive systems: Applications and payoffs*; Warminster, (may 1980).

[IBM, 1986]

Hardware Adaptations & Special Application Software For Persons with Disabilities; IBM Reference Manual, IBM National Support Center for Persons with Disabilities, (1986).

[IBM, 1987a]

Can I use an IBM computer if I can only press one key at a time?; IBM National Support Center for Persons with Disabilities, (January 1987).

[IBM, 1987b]

Can an IBM computer be operated with only a switch?; IBM National Support Center for Persons with Disabilities, (January 1987).

[IHM, 1988]

Interface Homme-Machine, notion d'ergonomie; Cours de seconde maîtrise en Informatique, FUNDP, Namur, (1988).

[Jakobson R., Fant C.G.M. & Halle M., 1951]

R. JAKOBSON, C.G.M. FANT & M. HALLE : *Preliminaries to Speech Amalysis, the Distinctive Features and Their Correlates*; MA:M.I.T. Press, Cambridge, (1951).

[J00ST, 1979]

M.G. JOOST & F.E. PETRY : *Design factor in th use of isolated word recognition for speech therapy*; Proceedings Human factors society, pp 30-34 (1979).

[KIERKEGAARD, 1850]

S.A. KIERKEGAARD : *La maladie à la mort*; trad. Tisseau, I, C, A, b, p33, (1813-1855).

[KLASS, 1982]

P.J. KLASS : *Technique benefits novice technicians*; Aviation Week and Space Technology, pp 133-135, (11 octobre 1982).

[KLEIN, 1974]

G. KLEIN : "Qu'est-ce qu'une machine ?" in *Histoire de machines*; Anthologie, Librairie Générale Française, Paris, (1974).

[Krasner, 1986]

J. KRASNER : *Case study : PC + voice card = overnight ordering*; Systems/3X word, Vol. 14, n°4, pp 44,68, (April 1986).

[KURZWEIL, 1986]

R. KURZWEIL : *Les secrets de Raymond Kurzweil*; Science & Vie Micro, n°28, pp 72-75, (Mai 1986).

[KURZWEIL-AI, 1988]

Correspondance avec KURZWEIL applied Intelligence, Inc.

[LANE, 1980]

N.E. LANE & S.D. HARRIS : *Conversations with weapon systems : Crewstation applications of interactive voice technology*; Warminster, (janvier 1980).

[LAZZARO, 1986]

J.J. LAZZARO : *Opening in the word place of the disabled*; High technology, Vol. 6, n°3, pp 54-56, (March 1986).

[LEA, 1980]

W.A. LEA (Editeur), *Trends in speech recognition*, Englewood Cliffs, NJ : Prentice-Hall. (1980).

[LEA, 1981]

W.A. LEA : *A research program to advance airborne uses of voice I/O*; Speech Science Publication, Santa Barbara, Calif. (octobre 1981).

[LEGGET-WILLIAMS, 1984]

J. LEGGET & G. WILLIAMS : *An empirical investigation of voice as an input modality for computer programming*; International Journal of Man-Machine Studies, Vol. 21, n°6, pp 493-520, (1984).

[LERMAN, 1980]

L. LERMAN : "Traceability via voice data entry in hybrid manufacturing at Lockheed Missile Systems Division"; in S. HARRIS (editor) : *Voice Interactive systems: Applications and payoffs*; Warminster, (may 1980).

[LEROY, 1988]

H. LEROY : *Compilation, matière approfondie*; Cours de seconde maîtrise en Informatique, FUNDP, Namur, (1988).

[Lesser V.R., 1975]

V.R. & A. LESSER : *Organization of the HEARSAY II Speech Understanding System*; IEEE Transaction on Acoustics, Speech and Signal Processing; 23, p 11-23, (1975).

[LOWERRE & REDDY, 1980]

B. LOWERRE & R. REDDY : "The Harpy Speech Understanding System" in W.A. Lea (Edition), *Trends in speech recognition*, Englewood Cliffs, NJ : Prentice-Hall. (1980).

[Lowerre B.T., 1976]

B.T. LOWERRE : "*The HARPY Speech Recognition System*", Technical Report; Carnegie-Mellon University, Dep. of Computer Science, (1976).

[MADLIN, 1986]

N. MADLIN : *Conversant computers; Management review*, Clifton, Vol. 75, n°4, pp 59-60, (April 1986).

[MANDEL, 1984]

A. MANDEL : *The integration of speech technology into a microelectric hybrid inspection application*; Proceedings American Voice I/O System Application (AVIOS) Conference, Arlington, Virginie, (11-13 sept. 1984).

[MARIANI, 1982]

J. MARIANI : "*The AESOP Continuous Speech Understanding System*", Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing 82; Paris, (1982).

[MARTIN-HERZET-JADOUL, 1987]

D. MARTIN, G. HERZET & P. JADOUL : *Cles pour PC et compatibles*; PSI édition, Paris, (1987).

[McCauley, 1984, pp. 131-166]

M.E. McCAULEY : "Human factors in voice technology".; in F.A. Muckler (editor), *Human Factors Review 1984*; The Human Factors Society, Santa Monica, Calif., pp 131-166, (1984).

[McCracken D.L., 1981]

D.L. MC CRACKEN : "A Production System Version of the HEARSAY II Speech Understanding System", University Research Press; (1981).

[McLuhan-Fiore, 1967]

M. MC LUHAN & Q. FIORE : *The Medium is the Message*; (1967).

[MERCIER, 1980]

G. MERCIER and al. : "The KEAL Speech Understanding System" in *Spoken Language Generation and Understanding*; J.C.Simon (éditeur), (1980).

[MERCIER, 78]

G. MERCIER, P. QUINTON & R. VIVES : *KEAL : Un système pour un dialogue oral avec une machine*; Congrès AFCET, Gif sur Yvette, 2, pp 304-314, (novembre 1978).

[MEUNIER, 1988]

Communication privée avec Christian Meunier, paraplégique, collaborateur bénévole aux projets Aidami.

[MEUNIER, 1988]

Communication privée avec Christian Meunier, paraplégique, collaborateur bénévole aux projets Aidami.

[MICHEL-BENOIT, 1988]

B. MICHEL & M. BENOIT : *PC, XT et AT : Maintenance et améliorations*; BCM-série professionnelle, Banneux, (1988).

[Michaelis-Wiggins, 1982, pp. 149-178]

P.R. MICHAELIS & R.H. WIGGINS : "A human factors engineer's introduction to speech synthesizers", in A. Brade & B. Schneiderman, *Directions in Human-Computer Interaction* ; Ablex Publishing Co., p 149-178, (1982).

[MOUNTFORD, 1980]

S.J. MOUNTFORD & R.A. NORTH : *Voice entry for reducing pilot workload*; Proceedings Human Factors Society - 24th Annual Meeting, pp 185-189 , (1983).

[MURRAY et al. 1983]

J.T. MURRAY, J. VAN PRAAG, D. GILFOIL : *Voice versus keyboard motion*; Proceedings Human Factors Society - 27th Annual Meeting, P 103, (1983).

[NAEGLE, 1985]

T. NAEGLE : *Speech technology leaves the realm of science fiction*; Electronics Week 58 (16), (1985).

[NEUMEISTER, 1985]

B. NEUMEISTER : *Tetravox : la voix au service de son maître*; Golden n°11, pp 22-26, (déc 84-jan 85).

[NORTH, 1982]

R. NORTH & W. LEA : *Application of advanced speech technology in manned penetration bombers*; Wright-Patterson AFB, Ohio, (Mars 1982).

[NYE, 1982]

J.M. NYE : *Human factors analysis of speech recognition systems*; Speech Technology, 1(2), pp 50-57; (1982).

[ODOR-SHARP, 1981]

J.P. ODOR & S. SHARP : *How microcomputers can help severely disabled people*; Proceedings, Disability and Technology in the 80's, pp 105-111, (Mars 1981).

[P.B. DENES et M.V. MATHEWS, 1960]

P.B. DENES & M.V. MATHEWS : "*Spoken digit recognition using time-frequency pattern matching*", J. Acoust. Soc. Amer.; Volume 32, p 1450-1455, (1960).

[PC-MAGAZINE, 1987]

Anonymous : *Voice-Recognitin System*; PC-Magazine, Ziff-Davis Publishing Company, (27 octobre 1987).

[PHILIP-YOUNG, 1987]

G. PHILIP & E.S. YOUNG : *Man machine interaction by voice : developments in speech tecnology (part 1 : the state of the art) and (part 2 : General applications, and potential applications in Libraries and Information Services)* Journal of informations science (principles and practice), Amsterdam, Vol. 13, n°1, pp 3-14, (1987).

[PIERREL, 1975]

J.M. PIERREL : *Contribution à la compréhension automatique du discours continu*, Thèse du troisième cycle, Université de Nancy I. (1975)

[PIERREL, 1987]

J.M. PIERREL : "Aspects of Man-Machine Voice Dialog"; in *Fundamentals in Computer Understanding: Speech and Vision*, J.P. Haton editor, Cambridge University Press, pp 249-274, (1987).

[PIERREL, 82]

J.M. PIERREL : *Use of linguistic Constraints for automatic continuous speech understanding : the MYRTILLE II system*, T.S.I., Vol. 1, n°5, PP 329-346, North Oxford Academic, (1982)

[POHL, 1968]

J. POHL : *Symboles et Langages*; Sodi, Paris et Bruxelles, T II, p10, (1968)

[PUBLIC-LAW-99-506, 1986]

Congress Public Law 99-506, "Section 508" : *Electronic Equipment Accessibility for Disabled Workers* :

(a)(1) The Secretary through the National Institute on Disability and Rehabilitation Research and the Administrator of the General Services, in consultation with the electronics industry, shall develop and establish guidelines for electronic equipment accessibility designed to insure that handicapped individuals may use electronic office equipment with or without special peripherals..."

[PVBA, 1988]

P. VAN BASTELAER : *Téléinformatique et réseaux*; Cours de seconde maîtrise en Informatique, FUNDP, Namur, (1988).

[QUINTON, 1982]

P. QUINTON : *Utilisation de contraintes syntaxiques pour la reconnaissance de la parole continue*; Technique et Sciences Informatiques, Vol. 1, n°3, pp 233-248, (1982).

[RADAR, 1988]

correspondance avec The Royal Association for Disability and Rehabilitation

[RATY, 1988]

Communication privée au sujet de considérations communes
Travaux en vue de l'obtention d'une M. Sc. Thésis (Finlande).

[REHM, 1980]

B.E. REHM : *Physical access control using automatic speaker verification*; Proceedings IEEE region 5 Conference, pp 40-44, (1980).

[Reddy D.R., L.D. Erman & R.D. Neely, 1973]

D.R. REDDY, L.D. ERMAN & R.D. NEELY : *"A Model for a System for Machine Recognition of Speech"*, IEEE Trans. A.U.; 21, pp 229-238, (1973).

[ROTHBERG, 1980]

M. ROTHBERG : *Applying automatic Speech-recognition to Data entry*; Mini-Micro Systems 13(11), pp 153-162, (1980).

[SAINT-SETIERS, 1985]

M. SAINT-SETIERS : *Tétraplégique, et pourtant 70% d'autonomie grâce à l'ordinateur*; Golden n°11, pp 28-29, (déc 84-jan 85).

[SCHNEIDERMAN, 1987]

SCHNEIDERMAN : *Designing the User Interface : Strategies for Effective Human-Computer Interaction*; Addison-Wesley Publishing Company, U.S.A., (1987).

[SCOTT, 1983]

B.L. SCOTT : *Voice recognition systems and strategies*; Computer Design, pp 67-70, (Jan. 1983)

[SCOTT, 1988]

Correspondance avec Scott Instruments Corporation

[Schmandt, 1985, pp 133-159]

C. SCHMANDT : *"Voice communication with computers"*, in H. Rex Hartson, (editor), *Advances in Human-Computer Interaction* ; Volume 1, Ablex Publishing Co., Norwood, New Jersey, p 133-159, (1985).

[SENIA, 1986]

A. SENIA : *Machines that talk are making noises in the factory*; Production, Vol. 97, n°2, pp 40-44, (Fev. 1986).

[SEYMOUR, 1986]

J. SEYMOUR : *Computers that "Talk" and "Listen" part 2- Speech recognition*; Today's office, Vol. 21, n°6, pp 10-14, (Nov. 1986).

[SHANNON-WEAVER, 1964]

C.E. SHANNON & W. WEAVER : *The mathematical theory of communication*; Urbana, University of Illinois Press, (1964).

[SHEPPECK, 1984]

R.A. SHEPPECK : *The application of a voice interactive computer in organ transplantation : the Natco 24-alert system*; Proceedings American Voice I/O System Application (AVIOS) Conference, Arlington, Virginie, (11-13 sept. 1984).

[SHUTOH et al, 1984]

T. SHUTOH, S. TSURUTA, R. KAWAI & SHUTOH M. : "Voice operation in CAD system, in Human Factors in *Organizational Design*, H.W. HENDRICK & O. BROWN (editors), Elsevier Science Publisher, A'dam, pp 205-209 (1984).

[SPEECH +, 1988]

Correspondance avec Speech Plus incorporated

[SPERIANDO & SCAPIN, 1987]

J.C. SPERANDIO & D.L. SCAPIN : "Ergonomics aspects of man-machine communications", in *Fundamentals in Computer Understanding: Speech and Vision*, J.P. Haton editor, Cambridge University Press, pp 80-90, (1987).

[STRIEB-STOKES, 1980]

M.I. STRIEB & J.M. STOKES : "Military applicaioin of task oriented grammars" in S. HARRIS (editor) : *Voice Interactive systems: Applications and payoffs*; Warminster, (may 1980).

[THIRIONET, 1987]

P. THIRIONET : *Conception d'interfaces standards pour une application destinée à des handicapés*; mémoire dn licence et maîtrise en informatique, FNDP, Namur, (septembre 1987).

[THOMAS-ROSSON, 1984]

J.C. THOMAS & M.B. ROSSON : *Human factors in synthetic speech*; Proceedings American Voice I/O System Application (AVIOS) Conference, Arlington, Virginie, (11-13 sept. 1984).

[TORTEL, 83]

C. TORTEL : *Des ordinateurs qui aident à vivre*; L'Ordinateur Individuel, n°47, pp 112-114, (Avril 1983).

[VAN HEMEL, 1980]

S.E. VAN HEMMEL, S.B. VAN HEMMEL, W.J. KING, R. BREAUX : *Training implications of airborne application of automated speech recognition technology*; Orlando, Floride, (1980).

[VAN PEURSEM, 1984]

R.S. VAN PEURSEM : *The individual industrial application of Voice I/O*; Proceedings American Voice I/O System Application (AVIOS) Conference, Arlington, Virginie, (11-13 sept. 1984).

[VESTERIG-PROPST, 1982]

R.E. VESTERIG & F.M. PROPST : *Speech input and video disk team up for military maintenance*; Speech Technology, 1(3), pp 73-75; (1982).

[VS-1000, 1986]

DRAGON SYSTEM inc : *VoiceScribe-1000, Speech Application Designer's Handbook*; release 2.10 (1986).

[VTT, 1986]

Annual report of the medical Engineering Laboratory; Valtion Teknillinen Tutkimuskeskus, Technical Research centre of Finland, Tampere (1986)

[WALLER, 1984]

L. WALLER : *Voice recognition boards for the IBM PC are making their way into the market place*; Electronics Week 57 (12), (1984).

[WEBER, 1985]

D.M. WEBER : *Kurzweil almost ready for the market*; Electronics Week 58 (16), (1985).

[WERKOWITZ, 1980]

E. WERKOWITZ : "Ergonomic considerations for the cockpit application of speech generation technology" in S. HARRIS (editor) : *Voice Interactive systems: Applications and payoffs*; Warminster, (may 1980).

[WICKE et al., 1983]

R. WICKE, R.A. ENGELHARDT, R. AWAD, L. LEIFER & M. VAN DER LOOS : *Evaluation of a speaker-dependant voice recognition unit as an input device for control of a robotic arm*; 6th Annual Conference on Rehabilitation Engineering, SanDiego, (juin 1983).

[WOLF J.J. & WOODS W.A., 1977]

J.J. WOLF & W.A. WOODS : "*The HWIM Speech Understanding System*", IEEE International Conference on Acoustics, Speech and Signal Processing 77; Hartford, (1977).

[YAMAGUCHI, 1983]

J. YAMAGUCHI : *International viewpoints*; Automotive Engineering, 91, pp 72-76, (1983).

[YUDIN et al., 1980]

M. YUDIN, G.H. SELL, T. REICH, M. CLAGNAZ, H.LOUIE & R. KOLWICZ : *A voice controlled powered wheelchair and environmental control for the severely disabled*; Medical Progress through Technology 7, pp 139-143, (1980).

ANNEXE A :
Liste des fabricants

Annexe A : Liste des fabricants

Nom	N°	Adresse	Localité	CP	Pays	Num-tél
ANTEX Electronics corp	16100	South Figueroa st.	Gardena	CA 90248	USA	(213)532-3092
Cherry Electrical Products corp.	3600	Sunset Ave	Waukegan	IL 60087	USA	(312)360-3500
Dragon Systems, Inc.	55	Chapel Street	Newton	MA 02158	USA	(617)965-5200
IBM Entry System Division	11400	Burnett Road	Austin	TX 78758	USA	(512)838-8308
Interstate Voice Products	1848	W. Sequoia Avenue	Orange	CA 92668	USA	(714)937-9010
ITT Information Systems	2350	Qume Drive	San Jose	CA 95131	USA	(408)945-8950
Key Tronic		P.O. Box 14687	Spokane	WA 99214	USA	(800)262-6006
Kurzweil Applied Intelligences	411	Waverly Oaks Road	Waltham	MA 02154	USA	(617)893-5151
MTI incorporated	14711	N.E. 29th PL., Suite 245	Bellevue	WA 98007	USA	(206)881-1789
NEC America, Inc.	8	Old Sod Farm Road	Melville	NY 11747	USA	(516)753-7361
POWER var	2691	Dow Avenue, Suite F	Tustin	CA 92680	USA	(714)544-9941
Roar Technology Inc	1230	Sheppard Ave. W, Unit 2	Downsview	ONT. M3K 1Z9	CANADA	(416)638-8916
Scott Intruments	1111	Willow Drive	Denton	TX 76201	USA	(817)387-9514
Speech Plus	461	N. Bernardo Ave.	Mountain View	CA 94039-7461	USA	(415)964-7023
Texas Instruments, Inc.		P.O. Box 53	Lubbock	TX 79408	USA	(800)842-2737
Voice Connection	17835	Skypark Circle, Suite C	Irvine	CA 92714	USA	(714)261-2366
Voice Recognition Systems	550	Battery Street, Suite 1716	San Francisco	CA 94111	USA	(415)788-2007
Votan	4487	Technology Drive	Fremont	CA 94538	USA	(415)490-7600
Words+, Inc.	1125	Stewart Court, Suite D	Sunnyvale	CA 94086	USA	(418)730-9588
NEC DEUTCHLAND	4	Klauserburgerstr.	MUNCHEN 80	D- 8000	GERMANY	00-49-89.93.0
NEC Corporation		33-1 SHIBA 5-CHOME	MINATO-KU TOKYO		JAPON	

ANNEXE B :
Liste des constructeurs
et produits associés

Table des matières	1
Annexe B : Liste des constructeurs et produits associés . .	2
introduction	2
Le hardware (machine, carte, ...) disponible	3
1 Cherry Electrical Products corp	3
2 Dragon Systems, Inc.	4
3 IBM Entry System Division	5
4 Interstate Voice Products	6
5 ITT Information Systems	7
6 Key Tronic	8
7 Kurzweil Applied Intelligences	9
8 MTI incorporated	10
9 NEC America Inc.	11
10 Roar Technology Inc	14
11 Scott Instruments	15
12 Voice Connection	16
13 Votan	17

Annexe B : Liste des constructeurs et produits associés

De nombreux constructeurs se sont attachés à construire des systèmes de reconnaissance vocale. La liste qui suit présente une partie des constructeurs contactés dans le cadre du mémoire, ceux qui ont répondu correctement à l'appel d'offre envoyée.

Dans la suite de la correspondance, peu d'entre eux ont rompu tout contact. Mais, il a été difficile de recevoir des renseignements techniques précis. Pour ne pas gêner le lecteur avec d'innombrables pages de détails inutiles, la liste qui suit est un résumé succinct par constructeur des renseignements à connaître sur le produit fabriqué.

Nous y trouverons l'adresse complète du fabricant, la dénomination exacte du produit, son prix, une brève description, le taux de reconnaissance (donné par le constructeur), la taille du vocabulaire, le nombre de locuteur, le type d'élocution, le temps de réponse, le type de sortie.

Une bonne partie des constructeurs dont il est fait mention dans la liste qui suit proviennent de la même source, un manuel de référence édité par IBM et qui présente entre autre de nombreux constructeur de système de reconnaissance vocale. Pour les personnes désireuses de plus d'informations au sujet de ce manuel, elles sont invitées à lire le troisième chapitre de ce mémoire, consacré aux handicapés faces à l'aide informatique et plus particulièrement à la reconnaissance vocale.

Chacun des produits cités est compatible avec la génération des ordinateurs de type Personal Computers de IBM ainsi qu'avec les logiciels s'exécutant sur ce genre de machine.

Le hardware (machine, carte, ...) disponible

1 Cherry Electrical Products corp

Nom: Cherry Electrical Products corp

Produit: la série Voice-Scribe distribuée en
collaboration avec Dragon Systems (cf
plus loin)

Adresse :

N°: 3600
Rue: Sunset Ave
Localité: Waukegan
CP: IL 60087
Pays: USA

Num-tél: (312)360-3500
Contact:

Importateur: Multiprox

2 Dragon Systems, Inc.

Nom: Dragon Systems, Inc.

Produit: Voice-Scribe 1000
Prix: \$ 995,00 version 2.10 remplacée en 1988
par le VoiceScribe 400
\$ 1600,00 et \$ 2500,00 pour la version
3.00
Description: Carte à insérer dans l'ordinateur,
software,...
Taux de reconnaissance: 99,5% aux tests
définis dans [DODDINGTON-SCHALK, 1981]
Taille du vocabulaire: 1000 mots maximum
par vocabulaire en mémoire
utilisation de la mémoire de l'ordinateur
Type de locuteur: monolocuteur
Type d'élocution: discrète, mot
entier
Temps de réponse:
Type de sortie: émulateur de
clavier
voir tests pour plus de détail

Produit: Voice-Scribe 20000
Prix: ? \$ 10.000,00
Description: même conception que le voicescribe 1000
mais reconnaissance de 20000 mots ainsi
que possibilité de dictaphone

Produit: DART-5000
Prix: \$ 5995,00
Description: station de travail (type PC XT) munie de
la reconnaissance vocale, tâche de
traitement de texte,
Taux de reconnaissance:
Taille du vocabulaire: 1000 mots
extensible à 5000
Type de locuteur: monolocuteur ?
Type d'élocution: discret
Temps de réponse:
Type de sortie: N/A

Adresse :

N°: 55
Rue: Chapel Street
Localité: Newton
CP: MA 02158
Pays: USA

Num-tél: (617)965-5200
Contact: J.M. Baker

Importateur: Multiprox
Contact: Willegems
Adresse 1: P.B. 71
Adresse 2: Kerkhoflaan, 14-16
CP: 9300
Localité: AALST
Pays: BELGIQUE
Num-tél: 053/787778

3 IBM Entry System Division

Nom: IBM Entry System Division

Produit: Voice Communications Option
Prix: 1295,00
Description: Carte à insérer dans l'ordinateur
Taux de reconnaissance:
Taille du vocabulaire:
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse:
Type de sortie: émulateur de
clavier, digitalisation de la voix,
interface téléphonique, synthèse vocale ?

Adresse :

N°: 11400
Rue: Burnett Road
Localité: Austin
CP: TX 78758
Pays: USA

Num-tél: (512)838-8308

Importateur: non disponible en Europe

4 Interstate Voice Products

Nom: Interstate Voice Products

Produit: VocaLink CSRB-LC
Prix: \$ 1250,00
Description: Carte à insérer dans l'ordinateur
Taux de reconnaissance: 99%
Taille du vocabulaire: 500 mots,
utilisation de la mémoire de l'ordinateur
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse:
Type de sortie: émulateur de
clavier

Produit: VocaLink serie 4000
Prix: \$ 5200,00
Description: Interface de reconnaissance (RS-232C)
Taux de reconnaissance: 99%
Taille du vocabulaire: 100 mots max 2
sec, stockables dans l'interface
Type de locuteur: monolocuteur
Type d'élocution: parole continue
Temps de réponse:
Type de sortie:

Adresse :

N°: 1848
Rue: W. Sequoia Avenue
Localité: Orange
CP: CA 92668
Pays: U A

Num-tél: (714)937-9010
Contact:

Importateur: ?

5 ITT Information Systems

Nom: ITT Information Systems

Produit: Voice Communication System
Prix: \$ 1350,00
Description: Carte à insérer dans l'ordinateur et
Interface téléphonique
(prise de message, prise de commande
téléphonique)
Taux de reconnaissance:
Taille du vocabulaire: 200 mots
Type de locuteur: plurilocuteur
(combinaison de monolocuteur)
Type d'élocution: discret
Temps de réponse:
Type de sortie: émulateur de
clavier, interface téléphonique

Adresse :

N°: 2350
Rue: Qume Drive
Localité: San Jose
CP: CA 95131
Pays: USA

Num-tél: (408)945-8950
Contact:

Importateur: ?

6 Key Tronic

Nom: Key Tronic

Produit: KB 5152 V
Prix: \$ 995,00
Description: clavier couplé à un système
d'introduction vocale pour PC (en
remplacement du clavier standard)
Taux de reconnaissance:
Taille du vocabulaire: max 9
vocabulaires en mémoire secondaire de 160
mots en mémoire-clavier de 1,25 sec max
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse:
Type de sortie: émulation de
clavier transmission des caractères par
le câble standard

Adresse :

N°:
Rue: P.O. Box 14687
Localité: Spokane
CP: WA 99214
Pays: USA

Num-tél: (800)262-6006
Contact: Att: Steve Quai

Importateur: Manudax
Contact: Ludo Ost
Adresse 1:
Adresse 2: Bld E. Bockstael, 5
CP: 1020
Localité: BRUXELLES
Pays: BELGIQUE
Num-tél: 02/4288780

7 Kurzweil Applied Intelligences

Nom: Kurzweil Applied Intelligences

Produit: Kurzweil Voicesystem (KVS)
Prix: \$ 6500,00
Description: Interface série (RS-232)
Taux de reconnaissance:
Taille du vocabulaire: 1000 mots par
vocabulaires, stockables dans l'interface
Type de locuteur: monolocuteur ou
plurilocuteur selon le vocabulaire
Type d'élocution: discret/continu
Temps de réponse:
Type de sortie: émulation de
clavier

Produit: Kurzweil Voiceterminal (KVT)
Prix: \$?
Description: Emulateur de terminal, station de travail
(type PC XT) munie de la reconnaissance
vocale
Taux de reconnaissance:
Taille du vocabulaire: 1000 mots max
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse:
Type de sortie: N/A

Produit: Kurzweil Voiceworks (KVV)
Prix: \$?
Description: "Dictaphone Automatique", station de
travail (type PC AT) munie de la
reconnaissance vocale pour des
applications de bureau, traitement de
texte, ...
Taux de reconnaissance:
Taille du vocabulaire: 20000 mots max
Type de locuteur:
Type d'élocution:
Temps de réponse:
Type de sortie: N/A

Adresse :

N°: 411
Rue: Waverly Oaks Road
Localité: Waltham
CP: MA 02154
Pays: USA

Num-tél: (617)893-5151
Contact: Att: Glenn Akers ou Robert Joseph

Importateur: USA

8 MTI incorporated

Nom: MTI incorporated

Produit: Pronounce Plus
Prix: 799,00
Description: carte à insérer dans l'ordinateur
Taux de reconnaissance:
Taille du vocabulaire: 256 mots maximum
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse:
Type de sortie: émulateur de
clavier, 256 caractères maximum par mot

Adresse :

N°: 14711
Rue: N.E. 29th PL., Suite 245
Localité: Bellevue
CP: WA 98007
Pays: USA

Num-tél: (206)881-1789
Contact: Att: Ellen Clark

Importateur: USA

9 NEC America Inc.

Nom: NEC America, Inc.

Produit: SAR-10 Voice Plus
"Speech Recognition and Audio Response"
Prix: \$ 599,00
Description: Carte à insérer dans l'ordinateur,
Taux de reconnaissance: 98%
Taille du vocabulaire: 250 mots de 2 sec
max stockables à la fois dans la carte
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse: .25 à .5 sec
Type de sortie: émulateur de
clavier, parole digitalisée (max 250
mots)

Produit: SR-10
"Speech Recognition"
Prix: \$?
Description: Interface de reconnaissance
Taux de reconnaissance: 98%
Taille du vocabulaire: 128 mots de 2 sec
max stockables à la fois dans l'interface
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse: .25 à .5 sec
Type de sortie: émulateur de
clavier, parole digitalisée (max 250
mots)

Produit: SR-100
"Speech Recognition"
Prix: \$ 2000
Description: Interface de reconnaissance (RS-232C,
ccitt V24, V28, ...)
Taux de reconnaissance: 99%+
Taille du vocabulaire: 120 mots de 2 sec
max stockables à la fois dans l'interface
Type de locuteur: monolocuteur (un
seul entraînement de la voix)
Type d'élocution: discret
Temps de réponse: 300 msec
Type de sortie: émulateur de
clavier

Produit: SR-2500
Prix: \$?
Description: Système intégré de reconnaissance et de
synthèse, réponse au téléphone (de 4 à 8
lignes téléphonique connectable) à relier
à un mainframe
Taux de reconnaissance: 99%+
Taille du vocabulaire: 16 à 20 mots
Type de locuteur: indépendant du
locuteur
Type d'élocution: discret
Temps de réponse: 128 msec
Type de sortie: voix synthétisée
et digitalisée (2048 sec/mots)

Produit: DP-200
Prix: \$ 7500,00
Description: Système intégré de reconnaissance vocale
continue
Taille du vocabulaire: 50 mots
option : 150 mots en mode continu, 500
mots en mode discret
Type de locuteur: monolocuteur
Type d'élocution: parole continue
ou reconnaissance discrète, une phrase
d'au plus cinq mots et de 4 sec maximum
Temps de réponse: 300 msec
Type de sortie: N/A

Adresse :

N°: 8
Rue: Old Sod Farm Road
Localité: Melville
CP: NY 11747
Pays: USA

Num-tél: (516)753-7361
Contact:

Importateur: Tritech via Deutschland ?
Contact: Gerard Coudron
Adresse 1:
Adresse 2: Chaussée de Ninove, 71
CP: 1750
Localité: SCHEPDAAL
Pays: BELGIQUE
Num-tél: 02/5690640

Autres contacts:

Nom: NEC Corporation
Adresse :
N°:
Rue: 33-1 SHIBA 5-CHOME
Localité: MINATO-KU TOKYO 108
CP:
Pays: JAPON

Nom: NEC DEUTCHLAND
Adresse :
N°: 4
Rue: Klauserburgerstr.
Localité: MUNCHEN 80
CP: D- 8000
Pays: GERMANY

10 Roar Technology Inc

Nom: Roar Technology Inc

Produit: VoiceKey
Prix: 49.900,00 F
Description: Taux de reconnaissance: 98%+
Taille du vocabulaire: 4 vocabulaires de
128 mots de 2 sec max
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse:
Type de sortie: émulateur de
clavier (64 caractères maximum)

Adresse :

N°: 1230
Rue: Sheppard Ave. W, Unit 2
Localité: Downsview
CP: ONT. M3K 1Z9
Pays: CANADA

Num-tél: (416)638-8916
Contact:

Importateur: FDS
Contact: Lecron
Adresse 1:
Adresse 2: Rue du Centre, 8
CP: 4800
Localité: VERVIERS
Pays: BELGIQUE
Num-tél: 087/228606

11 Scott Instruments

Nom: Scott Instruments

Produit: Corotechs VET3 Speech Recognition System
Description: Interface de reconnaissance (RS-232C)
Taux de reconnaissance: 99% (110 dB max)
Taille du vocabulaire: 200 mots de 3/4
sec max ou 100 de 1 sec max par
application
Type de locuteur: monolocuteur
Type d'élocution: parole continue
Temps de réponse: émulateur de
Type de sortie: clavier
voix digitalisée

Prix: \$ 8995,00

Adresse :

N°: 1111
Rue: Willow Drive
Localité: Denton
CP: TX 76201
Pays: UpA

Num-tél: (817)387-9514
Contact: Att: Janice Drake

Importateur: ?

12 Voice Connection

Nom: Voice Connection

Produit: HAL (Home Automation Link)
Prix: \$ 995,00
Description: système d'introduction vocale pour PC
couplé à un contrôle de l'environnement
Taux de reconnaissance:
Taille du vocabulaire: ?
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse:
Type de sortie: émulation de
clavier, synthèse vocale, télécommande,
... (en option)

Produit: Introvoice III
Prix: \$ 995,00
Description: clavier couplé à un système
d'introduction vocale pour PC (en
remplacement du clavier standard)
Taux de reconnaissance:
Taille du vocabulaire: 160 mots de 1.25
sec maximum
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse:
Type de sortie: émulation de
clavier par le câble standard (1 à 76
caractères maximum par "mot" enregistré)

Produit: Introvoice V
Prix: \$ 495,00
Description: Carte à insérer dans l'ordinateur,
software,...
Taux de reconnaissance:
Taille du vocabulaire: 400 mots de 1.25
sec maximum
utilisation de la mémoire de l'ordinateur
Type de locuteur: monolocuteur
Type d'élocution: discret
Temps de réponse:
Type de sortie: émulation de
clavier par le câble standard (1 à 1000
caractères maximum par "mot" enregistré
suivant l'extension mémoire utilisée)

Adresse :

N°: 17835
Rue: Skypark Circle, Suite C
Localité: Irvine
CP: CA 92714
Pays: USA

Num-tél: (714)261-2366
Contact:

Importateur: USA

13 Votan

Nom: Votan

Produit: Votan Voice Card
Prix: \$ 1200,00 - \$ 2700,00
Description: Carte à insérer dans l'ordinateur
Taux de reconnaissance: 99% (100dB)
Taille du vocabulaire: vocabulaires de 300 mots ou 425 mots
Type de locuteur: monolocuteur, plurilocuteur pour "1", "2", "3", "4", "Yes", "No" (en option)
Type d'élocution: Continu ou discret
Temps de réponse:
Type de sortie: émulateur de clavier (80 caractères maximum), digitalisation de la voix, interface téléphonique

Produit: VTR 6050 serie II
Prix: \$ 3500,00 - \$ 4800,00
Description: Interface de reconnaissance (RS-232C)
Selon configuration : terminal vocal (E/S vocal non programmable, maître) ou périphérique vocal (E/S vocal programmable, esclave)
Taux de reconnaissance: 99% (100dB)
Taille du vocabulaire: vocabulaires de 150 enregistrements de mots stockables dans l'interface (transfert entre mémoire hôte et mémoire interface : 2000 enregistrements maximum)
Type de locuteur: monolocuteur, plurilocuteur pour "1", "2", "3", "4", "Yes", "No" (en option)
Type d'élocution: Continu ou discret
Temps de réponse:
Type de sortie: émulateur de clavier (80 caractères maximum), digitalisation de la voix, interface téléphonique

Adresse :

N°: 4487
Rue: Technology Drive
Localité: Fremont
CP: CA 94538
Pays: USA

Num-tél: (415)490-7600
Contact: Att: Debra L. Wiggins

Importateur: ?

ANNEXE C :
Vocabulaire pour KB 5152 V
et analyse de proximité

Annexe C : Vocabulaire pour KB 5152 V et analyse de proximité

commande;<<3830B0B8>>
non;n
oui;o
aide-moi;{f1}
changer de formule;{f2}
changer de position;{f3}
changer de taille;{f4}
recalculer;{f5}
definir une zone;{f6}; curseur
transférer;{f7}
copier;{f8}
effet de zoom;{f9}
présenter autrement;{f10}
++quitter;{c d}q
+sauver;<<1D1C9C9D>>
+sauvetage;<<1D1C9C9D>>
changer;<<C646>>
annuler;{c m}a
menu;<<3830B0B8>>
A P P S;{c a}
divers;{c a}
disque;{c d}
cadre;{c c}
graphique;{c g}
rechercher;{c r}
modifier;{c m}
nombre;{c n}; menu
aligner a gauche;j
justifier a gauche;j
aligner a droite;a
justifier a droite;a
centrer;c
centrage;c
nombre de decimale;n
batir;{c b}; menu
cadre;c
table;t
feuille de calcul;f
base de donnees;b
macro;m
texte;{c t}; menu
souligner;s
italique;i
normal;n
gras;g
aligner a gauche;a
caler a droite;d
aligner a droite;d
justifier;j
centrer;c
marge gauche;m
limite droite;l
marge droite;l
renforcement de paragraphe;r
tabulation;t
edition;{c e}; menu
lancer;l
arreter;a
format;f

```
ejecter;e
suspendre;s
pause;p<<0181>>
controle;c
destination;d
option;o
qualite courrier;ffq<<0181>>
mode condense;ffm<<0181>>
chiffres;<<3830B0B8>>
1;1
2;2
3;3
4;4
5;5
6;6
7;7
8;8
9;9
0;0
curseur;<<3830B0B8>>
.en haut;<<48C8>>
.@monter;<<48C8>>
.en bas;<<50D0>>
.@descendre;<<50D0>>
.a gauche;<<4BCB>>
.@aller a gauche;<<4BCB>>
.a droite;<<4DCD>>
.@aller a droite;<<4DCD>>
.debut;<<47C7>>
.fin;<<4FCF>>
.fin de ligne;<<4FCF>>
.page en haut;<<49C9>>
.page en bas;<<51D1>>
.tout en haut;<<1D47C79D>>
.tout en bas;<<1D4FCF9D>>
:haut;<<48C8>>
:@up;<<48C8>>
:@monter;<<48C8>>
:bas;<<50D0>>
:@down;<<50D0>>
:@descendre;<<50D0>>
:gauche;<<4BCB>>
:@aller a gauche;<<4BCB>>
:droite;<<4DCD>>
:@aller a droite;<<4DCD>>
:debut;<<47C7>>
:fin;<<4FCF>>
:page en haut;<<49C9>>
:page en bas;<<51D1>>
:tout en haut;<<1D47C79D>>
:tout en bas;<<1D4FCF9D>>
:escape;<<0181>>
:break;<<1D46C69D>>
:majuscule;<<3ABA>>
:minuscule;<<3ABA>>
:ok;{r}
:entrer;<<3820A0B8>>
:sortir;<<381696B8>>
:niv sup;<<381696B8>>
:d-accord;{r}
:ca va;{r}
```


:c'est bon;{r}
:fermer;{r}
:ouvrir;{r}
:repertoire;{r}
:effacer;<<53D3>>
:detruire;<<53D3>>
:delete;<<53D3>>
:Backspace;{b}
:touchtab;{t}
:suivant;{t}
:*+++SILENCE
:FRAMEWORK;FW{R}

#	Word	Diff	#	Closest Word
000	commande . . .	16	026	nombre
001	non.	18	119	c'est bon
002	oui.	11	074	8
003	aide-moi . . .	17	050	limite droi
004	changer de f	18	005	changer de
005	changer de p	21	004	changer de
006	changer de t	18	015	+sauvetage
007	recalculer . .	26	033	nombre de d
008	définir une	18	054	édition
009	transférer . .	16	079	.@monter
010	copier	13	025	modifier
011	effet de zoo	30	083	.@aller a g
012	présenter au	18	119	c'est bon
013	++quitter. . .	27	079	.@monter
014	+sauver. . . .	22	060	pause
015	+sauvetage . .	19	006	changer de
016	changer. . . .	30	055	lancer
017	annuler. . . .	18	088	.fin de liq
018	menu	19	031	centrer
019	A P P S. . . .	27	037	+feuille de
----- press any key to continue -----				

#	Word	Diff	#	Closest Word
020	divers	31	073	7
021	disque	21	048	justifier
022	cadre.	08	035	cadre
023	graphique. . .	35	076	0
024	rechercher . .	29	033	nombre de d
025	modifier	15	010	copier
026	nombre.	22	052	renforcement
027	aligner a ga	07	045	aligner a g
028	justifier a	05	030	justifier a
029	aligner a dr	10	047	aligner a d
030	justifier a	14	128	suivant
031	centrer. . . .	20	055	lancer
032	centrage . . .	14	043	normal
033	nombre de de	24	010	copier
034	batir.	17	036	table
035	cadre.	10	022	cadre
036	table.	20	022	cadre
037	feuille de c	28	103	début
038	base de donn	16	036	table
039	macro.	20	089	.page en ha
----- press any key to continue -----				

#	Word	Diff	#	Closest Word
040	texte. . . .	21	124	detruire
041	souligner. .	21	002	oui
042	italique . .	22	086	.debut
043	normal . . .	13	032	centrage
044	gras	25	076	0
045	aligner a ga	12	027	aligner a g
046	decaler a dr	20	029	aligner a d
047	aligner a dr	08	029	aligner a d
048	justifier. .	25	125	delete
049	marge gauche	22	051	marge droit
050	limite droit	22	003	aide-moi
051	marge droite	21	049	marge gauch
052	renforcement	21	032	centrage
053	tabulation .	28	005	changer de
054	edition. . .	19	008	definir une
055	lancer . . .	16	031	centrer
056	arreter. . .	20	017	annuler
057	format . . .	11	080	.en bas
058	ejecter. . .	15	086	.debut
059	suspendre. .	16	098	@descendre

----- press any key to continue -----

#	Word	Diff	#	Closest Word
060	pause. . . .	06	093	haut
061	controle . .	13	060	pause
062	destination.	22	011	effet de zo
063	option . . .	23	033	nombre de d
064	qualite cour	25	103	debut
065	mode condens	19	033	nombre de d
066	chiffres . .	25	054	edition
067	1.	21	032	centrage
068	2.	18	119	c'est bon
069	3.	20	059	suspendre
070	4.	22	022	cadre
071	5.	18	035	cadre
072	6.	22	125	delete
073	7.	34	086	.debut
074	8.	11	002	oui
075	9.	30	038	base de don
076	0.	12	119	c'est bon
077	curseur. . .	17	127	touchtab
078	.en haut . .	10	093	haut
079	.@monter . .	11	095	@monter

----- press any key to continue -----

#	Word	Diff	#	Closest Word
080	.en bas. . .	15	069	3
081	.@descendre.	11	098	@descendre
082	.a gauche. .	32	075	9
083	.@aller a ga	08	045	aligner a g
084	.a droite. .	17	106	page en bas
085	.@aller a dr	06	102	@aller a dr
086	.debut . . .	11	103	debut
087	.fin	04	104	fin
088	.fin de lign	18	017	annuler
089	.page en hau	05	105	page en hau
090	.page en bas	08	106	page en bas
091	.tout en hau	06	107	tout en hau
092	.tout en bas	07	108	tout en bas
093	haut	08	060	pause
094	@up.	18	075	9
095	@monter. . .	08	079	.@monter
096	bas.	17	101	droite
097	@down. . . .	15	036	table
098	@descendre .	11	081	.@descendre
099	gauche . . .	29	060	pause
----- press any key to continue -----				

#	Word	Diff	#	Closest Word
100	@aller a gau	09	045	aligner a g
101	droite . . .	17	069	3
102	@aller a dro	08	085	.@aller a d
103	debut. . . .	09	086	.debut
104	fin.	05	087	.fin
105	page en haut	07	089	.page en ha
106	page en bas.	10	090	.page en ba
107	tout en haut	08	091	.tout en ha
108	tout en bas.	07	092	.tout en ba
109	escape . . .	17	126	Backspace
110	break. . . .	30	007	recalculer
111	majuscule. .	22	017	annuler
112	minuscule. .	15	086	.debut
113	ok	17	079	.@monter
114	entrer . . .	16	031	centrer
115	sortir . . .	26	010	copier
116	niv sup. . .	27	121	ouvrir
117	d-accord . .	25	003	aide-moi
118	ca va. . . .	21	026	nombre
119	c'est bon. .	11	076	0
----- press any key to continue -----				

#	Word	Diff	#	Closest Word
120	fermer . . .	28	076	0
121	ouvrir . . .	22	079	.@monter

122 repertoire .	18	003	aide-moi
123 effacer. . .	29	124	detruire
124 detruire . .	16	086	.debut
125 delete . . .	24	054	edition
126 Backspace. .	17	109	escape
127 touchtab . .	21	077	curseur
128 suivant. . .	17	030	justifier a
129 *+++SILENCE.	19	059	suspendre
130 FRAMEWORK. .	26	119	c'est bon

Press any key to continue

ANNEXE D :
Représentation graphique
de la voyelle 'a'

- Annexe D -

- 2 -

- 2 -

ANNEXE E :
Grammaire pour l'utilisation
du Dos avec VoiceScribe-1000
et automate correspondant

Grammaire pour l'utilisation du Dos avec VoiceScribe-1000

Grammaire

/*

File: DOS.LAN

DragonKEY Overlays for DOS.

Version 2.00 (C) Copyright Dragon Systems, Inc. 1985, 1986
Revised by B.Thiry

CONTENTS

- An Overview
- Always Active Words
- User-Definable Words
- Key Bindings
- Overlay Space Requirements

AN OVERVIEW

This section briefly describes the DOS Overlay. The best way to learn about the overlay is to try it out. Remember that you can use the Voice Menu to find out what words are active as you try different DOS commands.

With the DOS overlay, you can enter most DOS commands by voice. When you see the DOS prompt, you can say any of the following command names:

backup, change_directory, check_disk, compare, copy, date, delete,
directory, disk_compare, disk_copy, make_directory, print,
remove_directory, rename, restore, time, tree, type, volume

After you have spoken a command name, DragonKEY recognizes words corresponding to arguments for that command. For example, after you say "directory," you can say "pause" and "wide." These words generate the keystrokes "/P" and "/W" respectively.

The words "enter" and "cancel" generate the keystrokes ENTER and CTRL-C, respectively. Use these words, or press the corresponding keys to terminate a command. DragonKEY automatically resets the set of active words to the list of DOS commands. If the overlay and DOS ever get out of step, use "enter" or "cancel" to reset.

You can mix spoken and keyboard input. For example, you can say "print" and then type the name of the file to be printed. However, to be able to speak the arguments for a command, you must speak the command name. DragonKEY does not know to change the set of active words if you type the name.

You can enter drive designators by voice using words from the International Alphabet, as shown below. DragonKEY recognizes these words at the beginning of a command line, and in commands that accept filename arguments.

DRIVE SPOKEN WORD

a: alpha
b: bravo
c: charlie

ALWAYS ACTIVE WORDS

You can say the following words at any time:

WORD ACTION

DragonKEY Activates the DragonKEY Console.

menu Displays the Voice Menu.

scratch_that Erase the previously spoken command.

COMMANDS The COMMANDS group is a set of 5 user-definable words. If you add a word to this group, you can speak it at any time.

USER-DEFINABLE WORDS

The DOS Overlay contains the following user-definable word groups:

GROUP WORDS DESCRIPTION

ALL - This group contains the overlay's predefined words. By selecting the ALL group, you can change the name of a predefined word and edit its keystroke sequence.

The ALL group does NOT include slots for user-definable words. To add a word to the overlay, select one of the groups listed below.

FILENAMES 20 Use this group to add the names of files you frequently use. The words in the FILENAMES group are active for all commands that accept filenames as arguments.

MACROS 20 Use this group to define complete DOS commands you type frequently. To enter a command you have defined, say "macro" followed by the command name. After you say the command name, you can say the name of a file, from the FILENAMES group.

COMMANDS 5 The words in this group are always active. You can use this group to define complete DOS commands, as with the MACROS group. Alternatively, you can add words to output single keystrokes. For example, you could define the word "escape" to output the ESC key, which erases the current command line.

KEY BINDINGS

The following key bindings are defined in the DOS Overlay:

KEY	VCOM	ACTION
ALT-c	console	Activate the DragonKEY Console.
ALT-m	menu /permanent	Display the Voice Menu.
ALT-l	microphone on	Turn the microphone on.
ALT-o	microphone off	Turn the microphone off.
ALT-t	microphone momentary	Turn the microphone on for a single word.
ALT-r	remember	Start remembering a sequence of keystrokes.
ALT-d	define	Stop remembering a sequence of keystrokes, and enter the DragonKEY Console Define Word option.
ALT-b	backup backspace	Erase the previously spoken command.
enter	/echo reset	Both of these keys reset the overlay's grammar for a new command.
CTRL-c	/echo reset	

OVERLAY SPACE REQUIREMENTS

The table below estimates how much overlay space is required for this overlay. Estimates are shown both for the overlay alone, and for the overlay loaded together with the CONSOLE overlay.

The estimates assumes that you have trained all of the words in the overlay, so you may actually require less space. See the DragonKEY Manual for instructions on reserving overlay space.

OVERLAYS LOADED	SPACE ESTIMATE
DOS	45K bytes
DOS, CONSOLE	65K bytes

*****/

#define vcom /* Special key bindings for this overlay */

```
"bind a-c console
bind a-m menu /permanent
bind a-l microphone on
bind a-o microphone off
bind a-t microphone momentary
bind a-r remember
bind a-d define
bind a-b backup backspace
```



```
bind /echo enter reset
bind /echo c-c reset;
```

/* Below are the definitions for all of the words in the overlay's vocabulary, with the keystroke sequences generated by the words. The most-used words are first so they appear first in training.

```
*/
enter "enter";
cancel "c-c";
menu "a-m";
DragonKEY "a-c";
scratch_that "a-b";
macro;
alpha "a.";
bravo "b.";
charlie "c.";
copy "copy ";
date;
time "time'c-m";
delete "del ";
directory "dir ";
type "type ";
print "print ";
make_directory "mkdir ";
change_directory "chdir ";
remove_directory "rmdir ";
rename "rename ";
tree "tree ";
disk_copy "diskcopy ";
backup "backup ";
check_disk "chkdsk ";
compare "comp ";
disk_compare "diskcomp ";
```

```
execute;
```

```
subdirectories "/S";
modified "/M";
add "/A";
fix "/F";
verify "/V";
ascii "/A";
binary "/B";
zero "0";
one "1";
two "2";
three "3";
four "4";
five "5";
six "6";
seven "7";
eight "8";
nine "9";
dash "-.";
pause "/P";
wide "/W";
single_sided "/1";
eight_sectors "/8";
format "format ";
system "/S";
double_sided "/4";
cancel_mode "/C";
print_mode "/P";
terminate_mode "/T";
prompt "/P";
colon ".,.";
file "/F";
restore "restore ";
volume;
```

/* The root sentence definition

```
*/
_ROOT [menu DragonKEY scratch_that cmd1 cmd2 cmd3 cmd4 cmd5] =
(
  DR,
  macro (MACROS FIL*)#,
  backup ( FIL, subdirectories, modified, add, _BDAT)*,
  change_directory FIL*,
  check_disk ( FIL, fix, verify)*,
  compare FIL*,
  copy ( FIL, verify, ascii, binary)*,
  _DDAT ( DIG, dash)*,
  delete FIL*,
  directory ( FIL, pause, wide)*,
```

```

disk_compare ( _DR_S, single_sided, eight_sectors)*,
disk_copy ( _DR_S, single_sided)*,
execute _FIL,
format ( _DR, system, single_sided, eight_sectors, _FVOL,
double_sided)*,
volume,
make_directory _FIL*,
print ( _FIL, cancel_mode, print_mode, terminate_mode)*,
rename _FIL*,
restore ( _FIL, subdirectories, prompt)*,
remove_directory _FIL*,
time ( _DIG, colon)*,
tree ( _DR, file)*,
type _FIL*,
_WOL _DR*
) # (enter, cancel);

/* User-definable word groups
*/
* ALL =          /* All pre-defined words */

add, alpha, ascii, backup, binary, bravo, cancel, cancel_mode,
change_directory, charlie, check_disk, colon, compare, copy, dash, date,
delete, directory, disk_compare, disk_copy, double_sided, DragonKEY, eight,
eight_sectors, enter, execute, file, five, fix, format, four,
macro, make_directory, menu, modified, nine, one, pause, print,
print_mode, prompt, remove_directory, rename, restore,
scratch_that, seven, single_sided, six, subdirectories, system,
terminate_mode, three, time, tree, two, type, verify, volume, wide, zero;

* MACROS = macro1,macro2,macro3,macro4,macro5,macro6,macro7,macro8,macro9,
macro10,macro11,macro12,macro13,macro14,macro15,macro16,macro17,
macro18,macro19,macro20;

! macro1; ! macro2; ! macro3; ! macro4; ! macro5; ! macro6; ! macro7;
! macro8; ! macro9; ! macro10; ! macro11; ! macro12; ! macro13;
! macro14; ! macro15; ! macro16; ! macro17; ! macro18; ! macro19;
! macro20;

* FILENAMES = MODEM, LOTUS_123, DBASE, AUTRE_UTILISATEUR,
DOS, PCTEX_directory, CHERRY_directory,
file8,file9,file10,file11,file12,file13,file14,file15,file16,file17,
file18,file19,file20;

MODEM "MODEM.BAT"; LOTUS_123 "LOTUS_12.BAT"; DBASE "DBASE.BAT";
AUTRE_UTILISATEUR "AUTRE_UT.BAT"; DOS "DOS.BAT";
PCTEX_directory "C:\\pctex"; CHERRY_directory "C:\\rec-voc\\cherry";
! file8; ! file9; ! file10; ! file11; ! file12; ! file13;
! file14; ! file15; ! file16; ! file17; ! file18; ! file19; ! file20;

* COMMANDS = cmd1,cmd2,cmd3,cmd4,cmd5;

! cmd1; ! cmd2; ! cmd3; ! cmd4; ! cmd5;

/* Subproductions
*/
_DIG = zero,one,two,three,four,five,six,seven,eight,nine;
_FIL = _DR, FILENAMES;
_DR = alpha, bravo, charlie;
_DR_S " " = _DR;
_BDAT "/D" = date;
_DDAT "date'c-m'w" = date;
_FVOL "/V" = volume;
_WOL "vol " = volume;

Scores obtenus

Scoring Log File

Language File:    dos.lfd
Vocabulary File:  lab.voc
Number of Tokens: 1
Iterations:       1
Rejection Threshold: 30
Reset Statistics:  Y
Use Tape Device:  N
Adapt Model:      N
Distance Threshold: 80
Number of Symbols: 105
Production:       [All]
State:            ALL-STATES

```


Messages:

Error: (Confidence) Symbol -> Symbol Recognized
Close-Call: (Distance) Symbol -> Close-Call

Close-Call: (74) menu -> binary

Close-Call: (69) charlie -> file

Close-Call: (51) copy -> add
Close-Call: (49) date -> eight
Close-Call: (78) date -> delete

Error: (82) time -> nine
Close-Call: (47) time -> time
Close-Call: (48) time -> file

Close-Call: (7) delete -> eight
Close-Call: (80) delete -> tree

Close-Call: (29) directory -> ascii
Close-Call: (65) directory -> execute

Error: (59) print -> eight
Close-Call: (23) print -> print
Close-Call: (29) print -> three
Close-Call: (42) print -> type
Close-Call: (44) print -> menu
Close-Call: (47) print -> delete

Close-Call: (1) tree -> three

Close-Call: (2) backup -> enter
Close-Call: (46) backup -> date

Close-Call: (52) modified -> double_sided

Close-Call: (35) add -> colon
Close-Call: (44) add -> type
Close-Call: (70) add -> charlie
Close-Call: (77) add -> enter

Close-Call: (72) fix -> six
Close-Call: (74) fix -> enter

Close-Call: (8) verify -> restore
Close-Call: (26) verify -> double_sided
Close-Call: (38) verify -> volume
Close-Call: (48) verify -> print_mode

Close-Call: (78) binary -> enter

Error: (80) three -> tree
Close-Call: (3) three -> three

Close-Call: (39) four -> zero

Close-Call: (23) five -> file
Close-Call: (78) five -> charlie

Error: (94) six -> fix

Close-Call: (27) seven -> cancel
Close-Call: (29) seven -> modified
Close-Call: (40) seven -> colon
Close-Call: (49) seven -> system
Close-Call: (68) seven -> enter

Error: (80) eight -> date
Close-Call: (4) eight -> delete
Close-Call: (20) eight -> eight

Close-Call: (56) wide -> file

Close-Call: (51) colon -> add

Close-Call: (24) MODEM -> add

Statistiques obtenues

Statistics File:
Language: dos.ldf
Vocabulary: lab.voc

Correct: 62, Wrong: 5, Rejects: 0

Histograms:

Confidences:	Total:	Correct:	Wrong:	Rejects:
90-100	31	30	1	0
80-89	17	14	3	0
70-79	12	12	0	0
60-69	5	5	0	0
50-59	2	1	1	0
40-49	0	0	0	0
30-39	0	0	0	0
20-29	0	0	0	0
10-19	0	0	0	0

Close-Calls:

<10	6
<20	0
<30	9
<40	3
<50	11
<60	4
<70	3
<80	8
<90	10
<100	10

List of Errors:

Count, Symbol -> Symbol Recognized

1 time -> nine
1 print -> eight
1 three -> tree
1 six -> fix
1 eight -> date

List of Close-Calls: [Threshold = 80]

Count, Symbol -> Close-Call

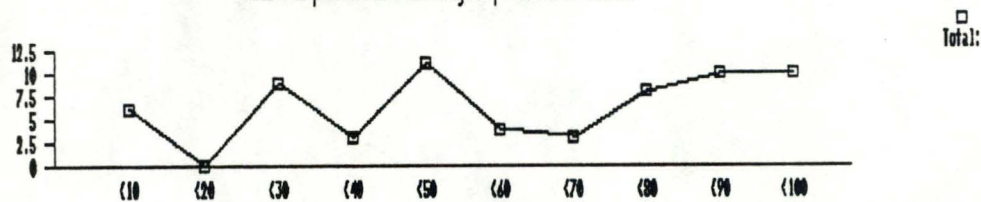
1 menu -> binary
1 charlie -> file
1 copy -> add
1 date -> delete
1 date -> eight
1 time -> time
1 time -> file
1 delete -> tree
1 delete -> eight
1 directory -> execute
1 directory -> ascii
1 print -> menu
1 print -> delete
1 print -> type
1 print -> print
1 print -> three
1 tree -> three
1 backup -> enter
1 backup -> date
1 modified -> double_sided
1 add -> enter
1 add -> charlie
1 add -> type
1 add -> colon
1 fix -> enter
1 fix -> six
1 verify -> double_sided
1 verify -> print_mode
1 verify -> restore
1 verify -> volume
1 binary -> enter
1 three -> three
1 four -> zero
1 five -> charlie
1 five -> file
1 seven -> enter
1 seven -> cancel
1 seven -> modified
1 seven -> system
1 seven -> colon
1 eight -> delete
1 eight -> eight
1 wide -> file
1 colon -> add
1 MODEM -> add

Histogramme

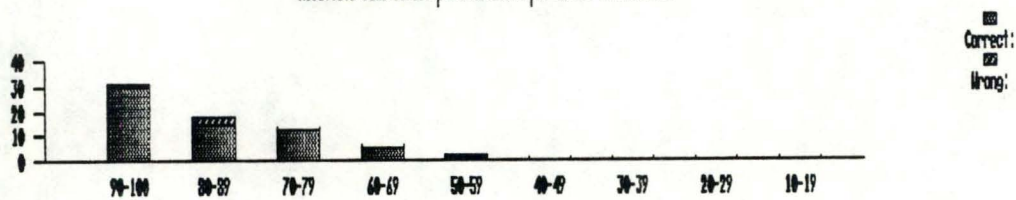
Confidences:	Total:	Correct:	Wrong:	Rejects:
90-100	31	30	1	0
80-89	17	14	3	0
70-79	12	12	0	0
60-69	5	5	0	0
50-59	2	1	1	0
40-49	0	0	0	0
30-39	0	0	0	0
20-29	0	0	0	0
10-19	0	0	0	0

Close-Calls:	
<10	6
<20	0
<30	9
<40	3
<50	11
<60	4
<70	3
<80	8
<90	10
<100	10

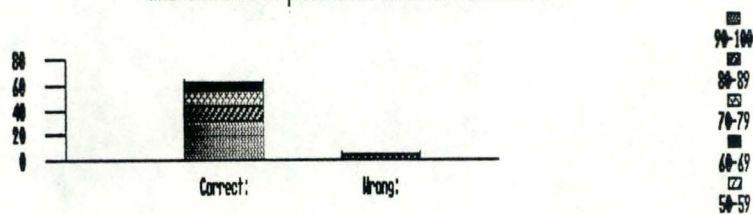
Nombre de prononciations non éloignées par niveau de confiance



Résultats cumulés des prononciations par niveau de confiance



Résultats cumulés des prononciations correctes et incorrectes



Parts des prononciations correctes par niveau de confiance



Correct:

ANNEXE F :
Alphabet et prononciation
française

Alphabet et prononciation française

Grammaire - alphabet

ALL = a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z;
a;b;c;d;e;f;g;h;i;j;k;l;m;n;o;p;q;r;s;t;u;v;w;x;y;z;

Scores obtenus - prononciation précise

Scoring Log File

Language File: alpha.ldf
Vocabulary File: alpha.voc
Number of Tokens: 1
Iterations: 1
Rejection Threshold: 1
Reset Statistics: Y
Use Tape Device: N
Adapt Model: N
Distance Threshold: 80
Number of Symbols: 26
Production: [All]
State: ALL-STATES

Messages:

Error: (Confidence) Symbol -> Symbol Recognized
Close-Call: (Distance) Symbol -> Close-Call

Close-Call: (14) b -> p
Close-Call: (27) b -> v
Close-Call: (36) b -> t
Close-Call: (40) b -> d

Close-Call: (42) c -> g

Error: (71) d -> t
Close-Call: (8) d -> v
Close-Call: (28) d -> p
Close-Call: (30) d -> d
Close-Call: (55) d -> b

Close-Call: (62) g -> j

Close-Call: (76) k -> a

Close-Call: (48) l -> z
Close-Call: (74) l -> m
Close-Call: (76) l -> y

Error: (86) n -> m
Close-Call: (15) n -> n

Close-Call: (28) p -> t
Close-Call: (33) p -> d
Close-Call: (54) p -> b

Close-Call: (21) q -> u
Close-Call: (64) q -> t

Close-Call: (59) r -> f

Close-Call: (12) t -> p

Close-Call: (71) u -> q

Close-Call: (70) v -> d
Close-Call: (80) v -> t

Statistiques obtenues - prononciation précise

Statistics File:

Language: alpha.ldf
Vocabulary: alpha.voc

Correct: 24, Wrong: 2, Rejects: 0

Histograms:

Confidences:	Total:	Correct:	Wrong:	Rejects:
90-100	16	16	0	0
80-89	6	5	1	0
70-79	3	2	1	0
60-69	1	1	0	0

50-59	0	0	0	0
40-49	0	0	0	0
30-39	0	0	0	0
20-29	0	0	0	0
10-19	0	0	0	0

Close-Calls:

<10	1
<20	3
<30	4
<40	3
<50	3
<60	3
<70	2
<80	5
<90	6
<100	7

List of Errors:

Count, Symbol -> Symbol Recognized

1 d -> t
1 n -> m

List of Close-Calls: [Threshold = 80]

Count, Symbol -> Close-Call

1 b -> d
1 b -> p
1 b -> t
1 b -> v
1 c -> g
1 d -> b
1 d -> d
1 d -> p
1 d -> v
1 g -> j
1 k -> a
1 l -> m
1 l -> y
1 l -> z
1 n -> n
1 p -> b
1 p -> d
1 p -> t
1 q -> t
1 q -> u
1 r -> f
1 t -> p
1 u -> q
1 v -> d
1 v -> t

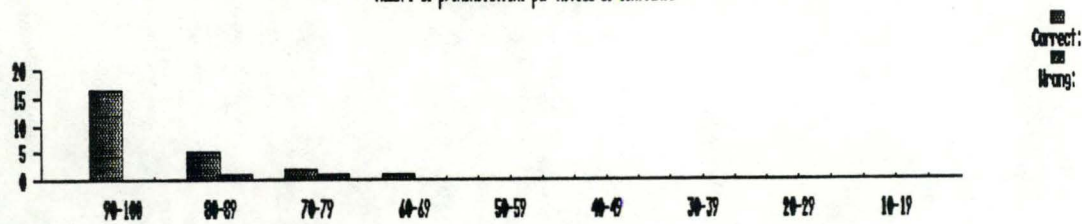
Histogramme

Confidences:	Total:	Correct:	Wrong:	Rejects:
90-100	16	16	0	0
80-89	6	5	1	0
70-79	3	2	1	0
60-69	1	1	0	0
50-59	0	0	0	0
40-49	0	0	0	0
30-39	0	0	0	0
20-29	0	0	0	0
10-19	0	0	0	0

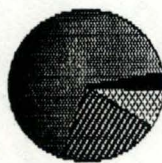
Close-Calls:

<10	1
<20	3
<30	4
<40	3
<50	3
<60	3
<70	2
<80	5
<90	6
<100	7

Nombre de prononciations par niveau de confiance



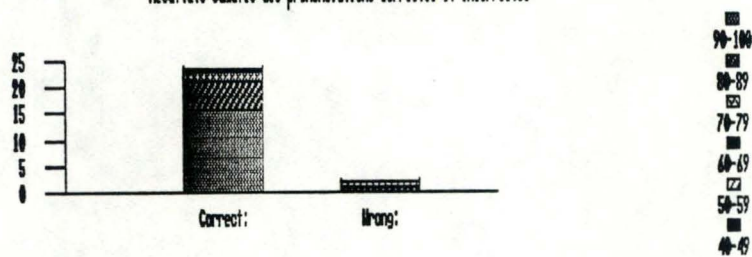
Parts des prononciations correctes par niveau de confiance



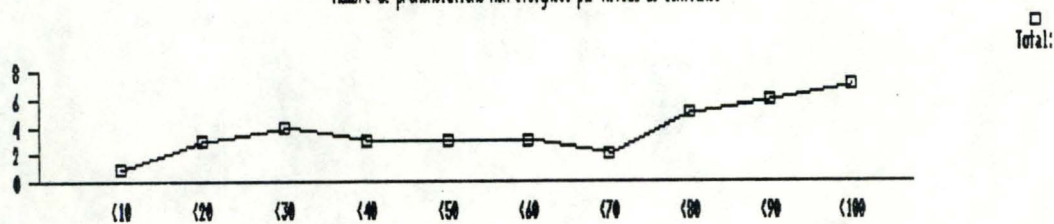
Correct:

90-100
80-89
70-79
60-69
50-59
40-49

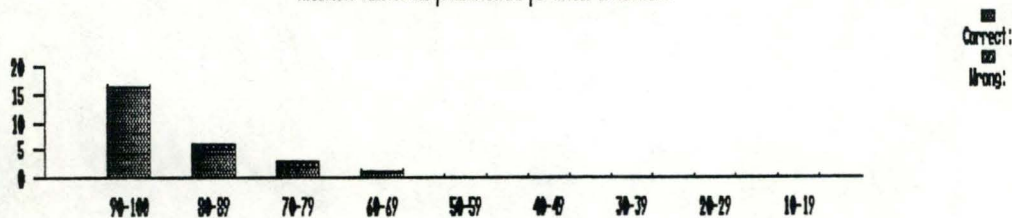
Résultats cumulés des prononciations correctes et incorrectes



Nombre de prononciations non éloignées par niveau de confiance



Résultats cumulés des prononciations par niveau de confiance



Scores obtenus - prononciation normale

Scoring Log File

Language File: alpha.ldf
 Vocabulary File: alpha.voc
 Number of Tokens: 1
 Iterations: 1
 Rejection Threshold: 1
 Reset Statistics: Y
 Use Tape Device: N
 Adapt Model: N
 Distance Threshold: 80
 Number of Symbols: 26
 Production: [All]
 State: ALL-STATES

Messages:
 Error: (Confidence) Symbol -> Symbol Recognized
 Close-Call: (Distance) Symbol -> Close-Call

Error: (76) b -> v
 Close-Call: (16) b -> t
 Close-Call: (23) b -> p
 Close-Call: (68) b -> d

Error: (73) d -> p
 Close-Call: (15) d -> t
 Close-Call: (20) d -> v
 Close-Call: (47) d -> d
 Close-Call: (52) d -> b
 Close-Call: (58) d -> g

Close-Call: (65) e -> k
 Close-Call: (68) e -> p

Close-Call: (44) f -> s
 Close-Call: (60) f -> m
 Close-Call: (76) f -> r
 Close-Call: (76) f -> y

Close-Call: (16) g -> v

Close-Call: (50) i -> u
 Close-Call: (53) i -> q
 Close-Call: (57) i -> j
 Close-Call: (61) i -> d
 Close-Call: (70) i -> b

Close-Call: (15) j -> g

Error: (75) l -> z
 Close-Call: (23) l -> m
 Close-Call: (24) l -> l
 Close-Call: (37) l -> r
 Close-Call: (66) l -> n

Error: (97) n -> m

Close-Call: (17) p -> d
 Close-Call: (22) p -> t
 Close-Call: (55) p -> b
 Close-Call: (58) p -> v

Error: (83) q -> u
 Close-Call: (7) q -> q

Error: (84) r -> m
 Close-Call: (54) r -> r
 Close-Call: (63) r -> f

Error: (84) t -> p
 Close-Call: (47) t -> t

Close-Call: (56) t -> d

Close-Call: (44) u -> q

Close-Call: (48) u -> d

Close-Call: (57) u -> t

Close-Call: (71) v -> p

Statistiques obtenues - prononciation normale

Statistics File:

Language: alpha.ldf

Vocabulary: alpha.voc

Correct: 19, Wrong: 7, Rejects: 0

Histograms:

Confidences:	Total:	Correct:	Wrong:	Rejects:
90-100	11	10	1	0
80-89	7	4	3	0
70-79	6	3	3	0
60-69	1	1	0	0
50-59	0	0	0	0
40-49	0	0	0	0
30-39	0	0	0	0
20-29	0	0	0	0
10-19	0	0	0	0

Close-Calls:

<10 1
 <20 5
 <30 5
 <40 1
 <50 5
 <60 10
 <70 7
 <80 4
 <90 2
 <100 6

List of Errors:

Count, Symbol -> Symbol Recognized

1 b -> v
 1 d -> p
 1 l -> z
 1 n -> m
 1 q -> u
 1 r -> m
 1 t -> p

List of Close-Calls: [Threshold = 80]

Count, Symbol -> Close-Call

1 b -> d
 1 b -> p
 1 b -> t
 1 d -> b
 1 d -> d
 1 d -> g
 1 d -> t
 1 d -> v
 1 e -> k
 1 e -> p
 1 f -> m
 1 f -> r
 1 f -> s
 1 f -> y
 1 g -> v
 1 i -> b
 1 i -> d
 1 i -> j
 1 i -> p
 1 i -> u
 1 i -> g
 1 i -> l
 1 i -> m
 1 i -> n
 1 i -> r
 1 p -> b
 1 p -> d
 1 p -> t
 1 p -> v
 1 q -> q

1 r -> f
1 r -> r
1 t -> d
1 t -> t
1 u -> d
1 u -> q
1 u -> t
1 v -> p

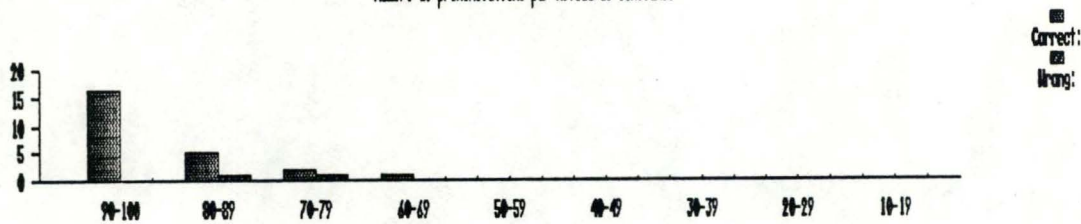
Histogramme

Confidences:	Total:	Correct:	Wrong:	Rejects:
90-100	11	10	1	0
80-89	7	4	3	0
70-79	6	3	3	0
60-69	1	1	0	0
50-59	0	0	0	0
40-49	0	0	0	0
30-39	0	0	0	0
20-29	0	0	0	0
10-19	0	0	0	0

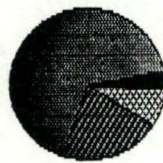
Close-Calls:

< 10	1
< 20	5
< 30	5
< 40	1
< 50	5
< 60	10
< 70	7
< 80	4
< 90	2
< 100	6

Nombre de prononciations par niveau de confiance



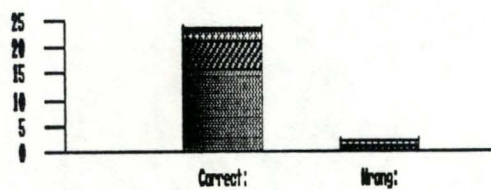
Parts des prononciations correctes par niveau de confiance



Correct:

90-100
80-89
70-79
60-69
50-59
40-49

Résultats cumulés des prononciations correctes et incorrectes

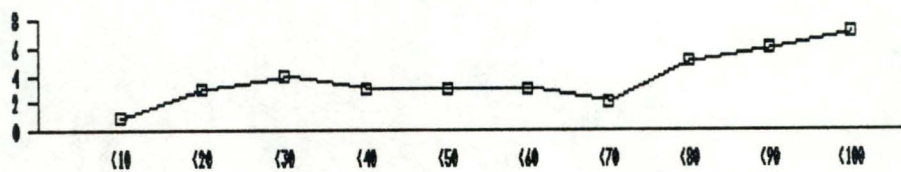


Correct:

Wrong:

90-100
80-89
70-79
60-69
50-59
40-49

Nombre de prononciations non éloignées par niveau de confiance



Total:

Résultats cumulés des prononciations par niveau de confiance



Correct:
Wrong:

ANNEXE G :
Analogies d'épellation
de l'alphabet latin

Analogies d'épellation de l'alphabet latin

<u>Lettre</u>	<u>International</u>	<u>U.S.A.</u>
A	Alfa	Adam
B	Bravo	Baker
C	Charlie	Charlie
D	Delta	David
E	Echo	Edward
F	Foxtrott	Frank
G	Golf	Georges
H	Hotel	Henry
I	India	Ida
J	Juliett	John
K	Kilo	King
L	Lima	Lewis
M	Mike	Mary
N	November	Nancy
O	Oscar	Otto
P	Papa	Peter
Q	Quebec	Queen
R	Romeo	Robert
S	Sierra	Susan
T	Tango	Thomas
U	Uniform	Union
V	Victor	Victor
W	Whisky	William
X	X-ray	X-ray
Y	Yankee	Young
Z	Zulu	Zebra

ANNEXES H, I, J

Programmes et grammaires développés dans le cadre du projet Aidami

Les textes des programmes et des grammaires développés dans le cadre du projet Aidami sont exclusivement disponibles auprès de monsieur Jean Ramaekers, Directeur de l'Institut.